



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Estimadores de Máxima Verossimilhança: Casos que não satisfazem as condições de regularidade.

Wellington Bernardo de Sousa

Orientador: Professor Peter Zörnig

Brasília

2018

Wellington Bernardo de Sousa

**Estimadores de Máxima Verossimilhança: Casos que não
satisfazem as condições de regularidade.**

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Professor Peter Zörnig

Brasília

2018

Wellington Bernardo de Sousa

Estimadores de Máxima Verossimilhança: Casos que não satisfazem as condições de regularidade. / Wellington Bernardo de Sousa. – Brasília, 2018-69 p. : il. (algumas color.) ; 30 cm.

Orientador: Professor Peter Zörnig

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 2018.

1. Estimadores de máxima verossimilhança. 2. Distribuição para grandes amostras e inferência com os E.M.V. 3. Casos que não satisfazem as condições de regularidade. 4. Aplicações.

Wellington Bernardo de Sousa

Estimadores de Máxima Verossimilhança: Casos que não satisfazem as condições de regularidade.

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Trabalho aprovado. Brasília, de junho de 2018:

Professor Peter Zörnig
Orientador

André Luiz Fernandes Cançado
Membro da Banca

Raul Yukihiro Matsushita
Membro da Banca

Brasília
2018

Resumo

Os estimadores de máxima verossimilhança são os estimadores que atribuem à amostra a maior probabilidade de ser observada. Estes estimadores possuem propriedades muito boas e sob as condições de regularidade pode-se aproximar a distribuição dos estimadores de máxima verossimilhança por meio da distribuição Normal e realizar inferência para os parâmetros populacionais através das estatísticas de Wald e da Razão de Verossimilhanças, por exemplo. No entanto, quando as condições de regularidade não estão satisfeitas, não é possível aproximar a distribuição desses estimadores por uma distribuição de probabilidades em particular e, portanto, há maior dificuldade para realizar inferência acerca dos parâmetros. Este trabalho trata dos estimadores de máxima verossimilhança com ênfase nos casos em que as chamadas condições de regularidade não estão satisfeitas, estas necessárias para aproximar a distribuição dos estimadores de máxima verossimilhança pela distribuição Normal, apresentando exemplos e algumas aplicações.

Palavras-chave: estimadores de máxima verossimilhança, condições de regularidade, distribuição dos estimadores de máxima verossimilhança, inferência para grandes amostras, distribuição Normal.

Abstract

The maximum likelihood estimators are the estimators that give the sample the highest probability of being observed. These estimators have very good properties and under regularity conditions one can approximate the distribution of the maximum likelihood estimators through the Normal distribution and perform inference for the population parameters through the Wald and 'Likelihood Ratio' statistics, for example. However, when the regularity conditions are not satisfied, it is not possible to approximate the distribution of these estimators by a particular probability distribution, and therefore, it is more difficult to make inference about the parameters. This work deals with the maximum likelihood estimators with emphasis in cases where the so-called regularity conditions are not satisfied, these are necessary to approximate the distribution of the maximum likelihood estimators by the Normal distribution, presenting examples and some applications.

Keywords: maximum-likelihood estimators, regularity conditions, MLE distribution, large sample inference, Normal distribution.

Lista de ilustrações

Figura 1	– Gráfico da função de verossimilhança $L(\lambda; x)$ em (1.3).	19
Figura 2	– Gráfico da função de log-verossimilhança $l(\lambda; x)$ em (1.4).	19
Figura 3	– Gráfico de $L(\theta; x)$, equação (1.14), para a amostra $\mathbf{X} = (0.2, 0.25, 0.5)$	22
Figura 4	– Gráfico de $l(\theta; x)$, equação (1.19), apenas em função de a , para a amostra $X^{(1)}$.	25
Figura 5	– Gráfico entre os valores preditos e resíduos do modelo (4.20).	57

Lista de tabelas

Tabela 1	– Valores da amostra $X^{(1)}$	24
Tabela 2	– Valores da amostra $X^{(2)}$	34
Tabela 3	– Valores da amostra $X^{(3)}$	39
Tabela 4	– Valores dos $b_{1j}^{*(1)}$ e $b_{1j}^{*(2)}$ para $j = 1, \dots, 15$.	40
Tabela 5	– Valores de $\psi^{(k)}$ em cada iteração $k = 0, 1, \dots, 15$	41
Tabela 6	– Valores dos $b_{1j}^{*(1)}$ e $b_{1j}^{*(2)}$ para $j = 1, \dots, 6$.	45
Tabela 7	– Número de mortes causadas por tuberculose e estimativa da população residente segundo a Unidade da Federação. Fonte: www.datasus.gov.br	49
Tabela 8	– Valores das estimativas nas 3 primeiras iterações.	51
Tabela 9	– Probabilidades estimadas de pertencer aos grupos 1 e 2 para cada Unidade da Federação.	52
Tabela 10	– Conjunto de dados Forbes	55
Tabela 11	– Valores preditos e resíduos do modelo (4.20).	56

Sumário

	Introdução	15
1	ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA	17
1.1	Função de Verossimilhança	17
1.2	Estimador de Máxima Verossimilhança	17
1.3	Exemplos do cálculo do E.M.V para alguns modelos	20
1.3.1	Modelo Poisson	20
1.3.2	Modelo Bernoulli	20
1.3.3	Modelo Normal	20
1.3.4	Casos não triviais	21
1.4	Propriedade de Invariância	26
2	DISTRIBUIÇÃO PARA GRANDES AMOSTRAS E INFERÊNCIA COM OS E.M.V.	27
2.1	Informação de Fisher	27
2.2	Distribuição para grandes amostras	29
2.3	Intervalos de Confiança e Testes de Hipóteses	30
2.3.1	Intervalos e Testes de Wald	31
2.3.2	Intervalo e Teste da Razão de Verossimilhanças	32
2.3.3	Inferência com Bootstrap	33
3	CASOS QUE NÃO SATISFAZEM AS CONDIÇÕES DE REGULARIDADE	35
3.1	Modelo Uniforme	35
3.2	Modelo Semi-Circular	36
3.3	Modelos de Mistura	36
3.3.1	Função de Verossimilhança com dados completos	37
3.3.2	Estimação de ψ com o Algoritmo E.M.	38
3.3.3	Mistura de Normais	42
3.3.4	Mistura de distribuições Poisson	43
4	APLICAÇÕES	47
4.1	Modelos de Mistura	47
4.1.1	Classificação do número de mortes por Tuberculose	47
4.1.2	Identificação de outliers	53
	Conclusão	59

APÊNDICE A - Códigos de R da subseção 4.1.1	61
APÊNDICE B - Códigos de R da subseção 4.1.2	65
REFERÊNCIAS	69

Introdução

Este texto foi escrito para ser submetido como o trabalho final de graduação em Estatística. O tema do trabalho é: Estimadores de Máxima Verossimilhança: Casos que não satisfazem as condições de regularidade. Tem como objetivo apresentar exemplos de estimadores de máxima verossimilhança sem os quais as correspondentes distribuições amostrais não podem ser aproximadas pela distribuição Normal e apresentar aplicações onde essa situação ocorre. As aplicações que ficamos mais interessados foram as escolhidas. Dessa forma, o trabalho visa evidenciar os casos em que as condições de regularidade não estão satisfeitas e apresentar a solução do Bootstrap para realizar inferência sem utilizar a distribuição Normal, e assim servir de exemplo para situações semelhantes.

O primeiro capítulo discorre a respeito dos estimadores de máxima verossimilhança de forma geral, apresentando alguns exemplos de como obtê-los analiticamente e numericamente. O segundo capítulo trata de forma breve as condições de regularidade, a distribuição assintótica dos estimadores de máxima verossimilhança e os principais métodos utilizados para realizar testes de hipóteses e construir intervalos de confiança com uso da distribuição Normal, os intervalos e testes de Wald e Razão de Verossimilhanças, restringido-se ao caso uniparamétrico. No terceiro capítulo são expostos alguns casos em que as condições de regularidade não são satisfeitas e no quarto capítulo são apresentadas aplicações onde há esse tipo de restrição. Nas aplicações foi utilizado o Bootstrap para aproximar a distribuição dos estimadores de máxima verossimilhança e realizar inferência.

Por fim, espera-se que o leitor se familiarize com os casos nos quais os estimadores de máxima verossimilhança não podem ser trabalhados com aproximações pela distribuição Normal, já que as condições de regularidade nem sempre estão satisfeitas.

1 Estimadores de Máxima Verossimilhança

O Método da Máxima Verossimilhança consiste basicamente em obter a estimativa mais verossímil dentro de uma amostra para o parâmetro populacional desconhecido. O valor mais verossímil que o estimador pode assumir é o mesmo que maximiza a probabilidade da amostra em questão ocorrer. Dessa forma, a função de probabilidade das variáveis aleatórias que representam uma amostra é encarada como função do correspondente parâmetro populacional de modo que possa ser obtida a estimativa desse parâmetro desconhecido. A próxima sessão dá mais ênfase a essa função denominada função de verossimilhança.

1.1 Função de Verossimilhança

Considerando uma amostra aleatória simples (X_1, X_2, \dots, X_n) da distribuição de uma variável aleatória X com função de densidade $f(x; \theta)$, em que θ é um escalar ou vetor, representando o parâmetro populacional desconhecido, denominamos a função $L(\theta; x)$ definida por

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta \quad (1.1)$$

como Função de Verossimilhança. Essa função depende do parâmetro populacional e os valores da amostra estão fixos. Θ é o espaço paramétrico de θ , o conjunto de valores que o parâmetro pode assumir.

O logaritmo natural da função de verossimilhança, denotada como $l(\theta; x)$, é chamada de Função de Log-Verossimilhança. Esta é mais conveniente de se trabalhar. O logaritmo é uma função crescente e, por isso, trabalhar com a função de log-verossimilhança é equivalente a trabalhar com a função de verossimilhança para efeitos de otimização. Assim, neste trabalho será utilizada a função de log-verossimilhança em vez da função de verossimilhança. A conveniência em utilizá-la para realizar cálculos reside no fato de que

$$l(\theta; x) = \log L(\theta; x) = \log \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \log f(x_i; \theta). \quad (1.2)$$

1.2 Estimador de Máxima Verossimilhança

O estimador de máxima verossimilhança de um parâmetro θ é o valor mais verossímil que θ pode assumir dado o que foi observado na amostra, ou seja, é o valor que maximiza a probabilidade de ocorrer a amostra observada. Assim, definimos o estimador de máxima verossimilhança da seguinte forma:

Definição 1.2.1 *O estimador de máxima verossimilhança de θ (E.M.V.) é qualquer valor $\hat{\theta}$ tal que $L(\hat{\theta}; x)$ é máxima.*

Portanto, $L(\hat{\theta}; x) \geq L(\theta; x)$ para todo $\theta \in \Theta$.

O valor do parâmetro que maximiza a função de verossimilhança é o mesmo que maximiza a função de log-verossimilhança, pois o logaritmo é uma função crescente. Este valor pode ser obtido identificando o ponto em que as primeiras derivadas se anulam e verificando se a matriz Hessiana é definida negativa neste ponto. Geralmente, a função de log-verossimilhança é côncava, logo o ponto onde a primeira derivada se anula é ponto de máximo, veja no Exemplo 1.1. Quando houver mais de um estimador de máxima verossimilhança ou quando ele não puder ser obtido através da primeira derivada, a função de verossimilhança deve ser estudada detalhadamente para encontrar o ponto de máximo.

Exemplo 1.1 *Modelo Poisson*

Seja $\mathbf{X} = (5, 4, 4, 4, 5)$ uma amostra aleatória simples de $X \sim \text{Poisson}(\lambda)$, $x = 0, 1, 2, \dots$ e $\lambda > 0$. Temos que

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \Rightarrow \quad L(\lambda; x) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad (1.3)$$

e

$$l(\lambda; x) = -n\lambda + \log(\lambda) \sum_{i=1}^n x_i - \sum_{i=1}^n \log(x_i!) \quad . \quad (1.4)$$

Com base na amostra do exemplo, temos que

$$L(\lambda; x) = \frac{e^{-5\lambda} \lambda^{22}}{(4!)^3 (5!)^2}$$

e

$$l(\lambda; x) = -5\lambda + 22 \log \lambda - (5!)^2 (4!)^3 \quad .$$

As figuras 1 e 2 mostram os gráficos das funções de verossimilhança e log-verossimilhança correspondentes à amostra do exemplo. Veja que a função de verossimilhança não é côncava, mas a função de log-verossimilhança é.

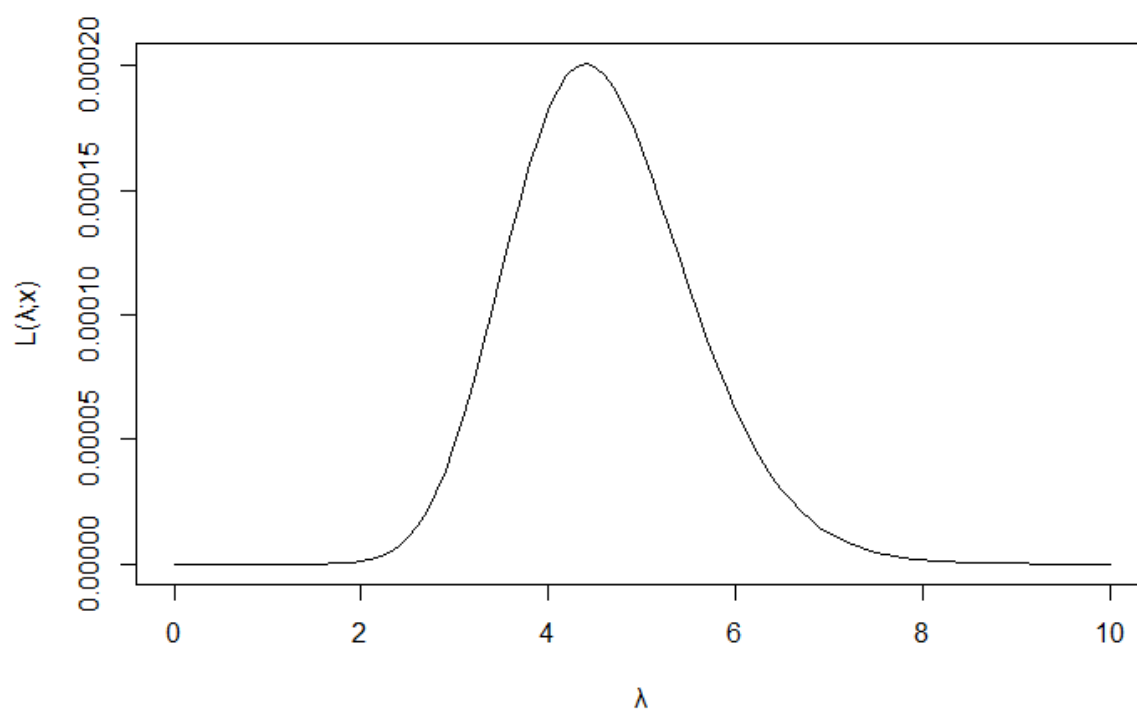


Figura 1 – Gráfico da função de verossimilhança $L(\lambda; x)$ em (1.3).

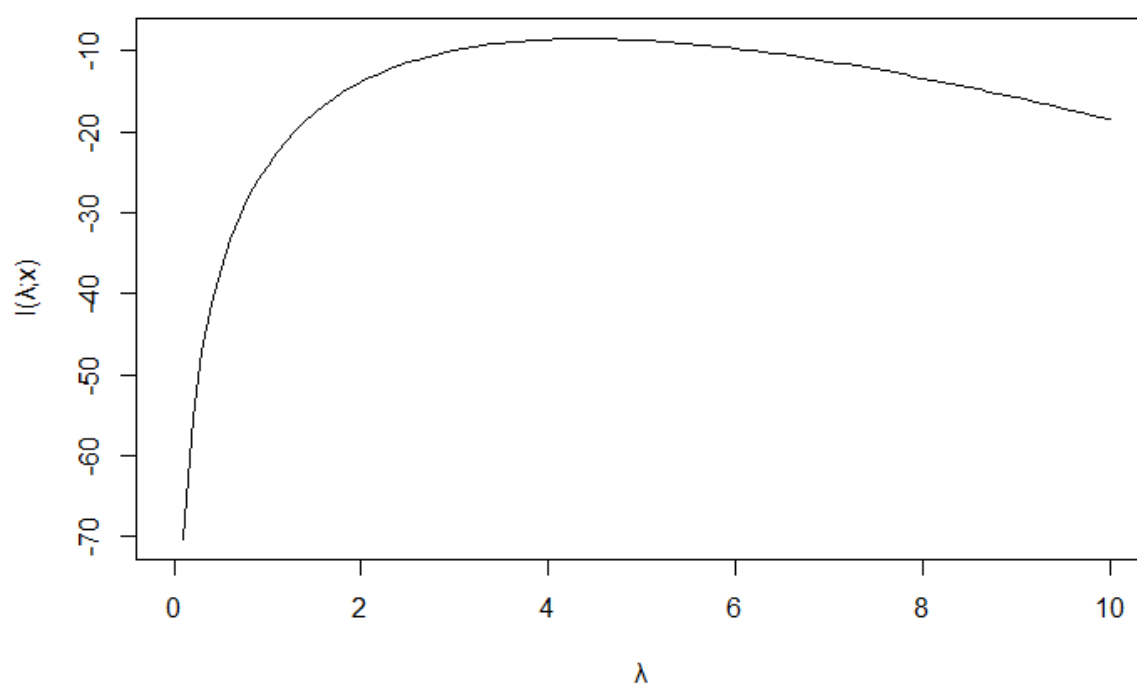


Figura 2 – Gráfico da função de log-verossimilhança $l(\lambda; x)$ em (1.4).

1.3 Exemplos do cálculo do E.M.V para alguns modelos

1.3.1 Modelo Poisson

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória simples da distribuição $Poisson(\lambda)$. De (1.4), temos que a primeira derivada de $l(\lambda; x)$ é da forma

$$\frac{\partial}{\partial \lambda} l(\lambda; x) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i \quad (1.5)$$

Igualando (1.5) a 0 e executando as operações algébricas necessárias, temos que $\lambda = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ é raiz de (1.5). Portanto \bar{x} , a média amostral, é o E.M.V. de λ .

A seguir, os E.M.V. serão escritos em função da variável aleatória com letra minúscula, pois representam o valor dentro de uma amostra em particular. Entretanto, serão utilizadas letras maiúsculas quando forem realizados cálculos referentes a variáveis aleatórias, como, por exemplo, o cálculo do valor esperado ou o da variância.

1.3.2 Modelo Bernoulli

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória simples da distribuição $Bernoulli(p)$. Sua função de probabilidade é dada por

$$f(x; p) = p^x (1 - p)^{1-x}; \quad x = 0, 1; \quad 0 < p < 1. \quad (1.6)$$

Da equação (1.2) temos que

$$l(p; x) = \sum_{i=1}^n x_i \log p + \left(n - \sum_{i=1}^n x_i \right) \log (1 - p) \quad (1.7)$$

e igualando (1.7) a 0, conclui-se que $\hat{p} = \bar{x}$ é o E.M.V. de p .

1.3.3 Modelo Normal

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória simples da distribuição $Normal(\mu; \sigma^2)$ e suponha que $\sigma = \sigma_0$ esteja fixo, $\sigma_0 > 0$. Queremos estimar μ . A densidade da distribuição Normal em questão é

$$f(x; \mu) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2} (x - \mu)^2}, \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}. \quad (1.8)$$

Logo, a sua função de log-verossimilhança é dada por

$$l(\mu; x) = -n(\log \sqrt{2\pi} + \log \sigma_0) - \frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (1.9)$$

Derivar a equação (1.9) com relação a μ , resulta na seguinte expressão

$$\frac{\partial}{\partial \mu} l(\mu; x) = \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \mu). \quad (1.10)$$

Igualando a última expressão a 0, temos que $\hat{\mu} = \bar{x}$ é o E.M.V. de μ .

Vale salientar, que nem sempre \bar{x} será o estimador de máxima verossimilhança.

1.3.4 Casos não triviais

Na maioria dos problemas de estimação por máxima verossimilhança, há mais de um parâmetro populacional a ser estimado e não é possível escrever o E.M.V. em termos dos valores observados na amostra, isto é, obter uma expressão analítica para o estimador. Neste caso, faz-se uso de algoritmos de otimização para encontrar aproximações razoáveis do valor do E.M.V. Em casos especiais não é possível encontrar o E.M.V. através da primeira derivada de $l(\theta; x)$. Este é o caso de uma distribuição Uniforme. Veja no exemplo a seguir.

Exemplo 1.2 Modelo Uniforme

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória simples da distribuição $Uniforme(0, \theta)$. Sua função de densidade é dada por

$$f(x; \theta) = \frac{1}{\theta}, \quad 0 < x < \theta, \quad \theta > 0. \quad (1.11)$$

A função de log-verossimilhança é

$$l(\theta; x) = -n \log \theta \quad (1.12)$$

e sua primeira derivada é

$$\frac{\partial}{\partial \theta} l(\theta; x) = -\frac{n}{\theta}, \quad (1.13)$$

que não possui raiz. Dessa forma, é necessário utilizar a função de verossimilhança e examiná-la minuciosamente.

Como $0 < x < \theta$, vale para todos os valores da amostra a seguinte função de verossimilhança:

$$L(\theta; x) = \begin{cases} 0, & \text{se } \theta < \max \{x_1, \dots, x_n\} \\ \frac{1}{\theta^n}, & \text{se } \theta \geq \max \{x_1, \dots, x_n\} \end{cases} \quad (1.14)$$

Com isso, o maior valor que $L(\theta; x)$ pode assumir será quando $\theta = x_{(n)} = \max \{x_1, \dots, x_n\}$. Veja na Figura 3 o gráfico de $L(\theta; x)$ considerando a amostra $(0.2, 0.25, 0.5)$ da distribuição Uniforme, o máximo da amostra é 0.5. Portanto, $\hat{\theta} = x_{(n)}$ é o E.M.V. de θ .

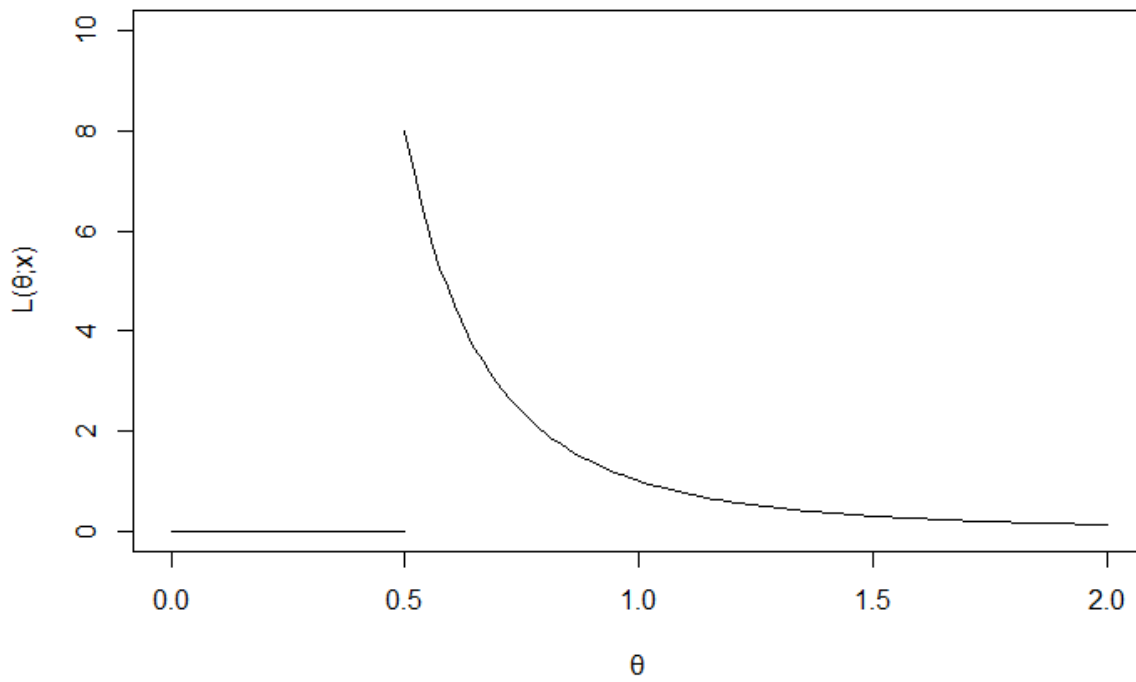


Figura 3 – Gráfico de $L(\theta; x)$, equação (1.14), para a amostra $\mathbf{X} = (0.2, 0.25, 0.5)$

Exemplo 1.3 *Modelo Normal (continuação)*

Considere a situação da subseção 1.3.3, mas agora admita σ desconhecido. Temos um vetor de parâmetros $\theta = (\mu, \sigma)$ e queremos estimar μ e σ . Como há 2 parâmetros populacionais desconhecidos, $L(\theta; x) = L(\mu, \sigma; x)$ é dada por

$$l(\theta; x) = -n(\log \sqrt{2\pi} + \log \sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2, \quad (1.15)$$

substituindo apenas σ_0 por σ qualquer, $\sigma > 0$. O ponto de máximo de $l(\theta; x)$ é o vetor $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$. Determinamos esse valor obtendo as derivadas parciais de $l(\theta; x)$ com relação a μ e σ e identificando os valores de μ e σ em que as duas derivadas parciais se anulam. Uma dessas tarefas já foi realizada anteriormente. Dessa forma, basta encontrar o valor de σ que maximiza $l(\theta; x)$, pois já temos $\hat{\mu} = \bar{x}$ que não depende de σ . Derivando $l(\theta; x)$ com relação a σ , temos que

$$\frac{\partial}{\partial \sigma} l(\theta; x) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2. \quad (1.16)$$

Igualando a última equação a 0, obtemos a raiz $\sigma = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu)^2}{n}}$. Substituindo μ por $\hat{\mu} = \bar{x}$, obtemos o estimador viciado $\hat{\sigma} = \sqrt{\left(\frac{n-1}{n}\right)s} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$. $\hat{\sigma}$ é proporcional à variância amostral $s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, e o vetor $\hat{\theta} = \left(\bar{x}, \sqrt{\left(\frac{n-1}{n}\right)s}\right)$ é o E.M.V. de $\theta = (\mu, \sigma)$.

Exemplo 1.4 *Modelo Beta*

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória simples da distribuição $Beta(a, b)$. Temos o vetor de parâmetros populacionais desconhecidos $\theta = (a, b)$. A densidade da distribuição Beta é da forma

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1, \quad a > 0, \quad b > 0, \quad (1.17)$$

onde $B(a, b)$ é a função Beta, dada por $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ e $\Gamma(x)$ é a função Gama. A função de log-verossimilhança de θ é dada por

$$l(\theta; x) = -n \log B(a, b) + (a-1) \sum_{i=1}^n \log x_i + (b-1) \sum_{i=1}^n \log(1-x_i) \quad (1.18)$$

e as derivadas parciais de $l(\theta; x)$ com relação a a e b são, respectivamente,

$$\frac{\partial}{\partial a} l(\theta; x) = n[\psi^{(1)}(a) + \psi^{(1)}(a+b)] + \sum_{i=1}^n \log x_i \quad e \quad (1.19)$$

$$\frac{\partial}{\partial b} l(\theta; x) = n[\psi^{(1)}(b) + \psi^{(1)}(a+b)] + \sum_{i=1}^n \log x_i \quad , \quad (1.20)$$

onde $\psi^{(n)}(x)$ é a $(n+1)$ -ésima derivada de $\log \Gamma(x)$, $\frac{d^{n+1}}{dx^{n+1}} \log \Gamma(x)$, denominada por função Poligama.

Como pode ser visto, não é possível explicitar analiticamente a raiz do sistema composto de (1.19) e (1.20). É necessário utilizar métodos de otimização para aproximar o valor da estimativa de máxima verossimilhança, como por exemplo o Método de Newton-Raphson. O uso de softwares computacionais é muito útil nestes casos, pois existe uma grande variedade de algoritmos de otimização implementados neles. Com base na amostra $X^{(1)}$ obtemos as estimativas $\hat{a} = 1,0643663$ e $\hat{b} = 0,5621073$. Os valores da amostra são apresentados na Tabela 1,

Tabela 1 – Valores da amostra $X^{(1)}$

0,99	0,46	0,86	0,71	0,16
------	------	------	------	------

Os valores 3 e 5 foram escolhidos arbitrariamente para iniciar o procedimento iterativo em função de a e b , respectivamente. Podemos utilizar gráficos e ou valores de outros estimadores para encontrar valores iniciais de forma que o algoritmo convirja para o valor correto da estimativa de máxima verossimilhança.

Exemplo 1.5 Modelo Beta (continuação)

Considere a situação do exemplo anterior, obter aproximações para os E.M.V. de a e b . O valor esperado de $X \sim \text{Beta}(a, b)$ é

$$E(X) = \frac{a}{a+b} \quad .$$

Assim, podemos utilizar a estimativa de momentos do valor esperado de X , a média amostral \bar{x} , para escrever \hat{b} em função de \hat{a} ,

$$\bar{x} = \frac{\hat{a}}{\hat{a} + \hat{b}} \quad \Rightarrow \quad \hat{b} = \frac{\hat{a}(1 - \bar{x})}{\bar{x}} \quad ,$$

Isso possibilita desenhar o gráfico de $l(\theta; x)$ apenas em função de \hat{a} e ver qual valor de a maximiza a log-verossimilhança. Em seguida, buscamos valores para melhorar as estimativas de a e b . Para a amostra do exemplo anterior, $\bar{x} = 0,6359944$ e o gráfico da função de log-verossimilhança é a apresentado na Figura 4.

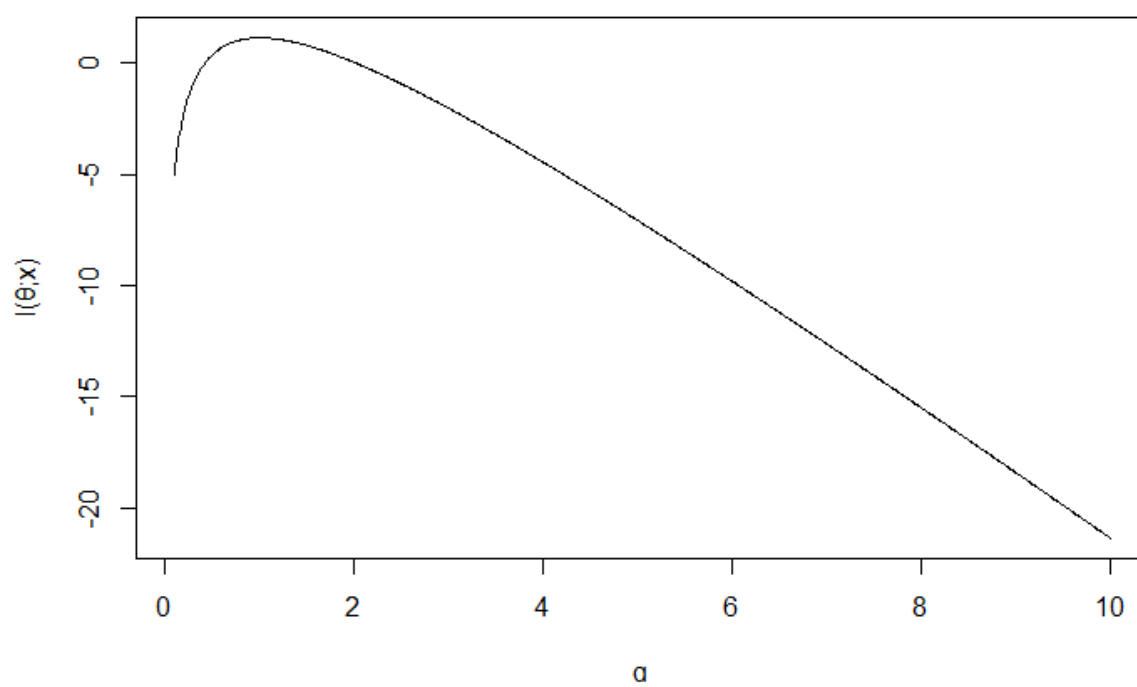


Figura 4 – Gráfico de $l(\theta; x)$, equação (1.19), apenas em função de a , para a amostra $X^{(1)}$.

Vemos que o valor em que $l(\theta; x)$ é maior ocorre em torno de $a = 1$, portanto, podemos iniciar o procedimento iterativo do exemplo anterior com $\hat{a} = 1$ e $\hat{b} = \frac{1-0,6359944}{0,6359944} = 0,5723409$ o qual retorna as mesmas estimativas de a e b obtidas no exemplo anterior.

$$\hat{a} = 1,0643663 \quad e \quad \hat{b} = 0,5621073 \quad .$$

1.4 Propriedade de Invariância

A Propriedade de Invariância permite calcular o estimador de máxima verossimilhança de uma função do parâmetro populacional, desde que a função real seja invertível (BOLFARINE; SANDOVAL, 2010). Se $\hat{\theta}$ é o E.M.V. de θ e g é uma função real, então

$$\widehat{g(\theta)} = g(\hat{\theta}) \quad (1.21)$$

é o estimador de máxima verossimilhança de $g(\theta)$.

Exemplo 1.6 Variância do Modelo Normal

Considere o problema da subseção 1.3.3, $s = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$ é o E.M.V. de σ . Agora queremos ueremos estimar σ^2 , a variância populacional de X . Então, pela Propriedade de Invariância,

$$\widehat{\sigma^2} = \hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} \quad (1.22)$$

é o E.M.V. de σ^2 .

Para mais exemplos e um texto mais aprofundado sobre o tema de estimadores de máxima verossimilhança, consulte (RATHIE; ZÖRNIG, 2012) e (BUSSAB; MORETTIN, 2013).

2 Distribuição para grandes amostras e inferência com os E.M.V.

Os resultados apresentados a seguir serão enunciados para o caso uniparmétrico, mas podem ser estendidos para o caso multiparmétrico de estimadores de máxima verossimilhança.

Seja $\hat{\theta}$ o estimador de máxima verossimilhança de θ como na definição 1.2.1. Sob as condições de regularidade, veja (MILLER, 2011), a distribuição de $\hat{\theta}$ converge em distribuição para uma Normal, a ser especificada posteriormente. Vamos agora enunciar alguns conceitos necessários para estabelecer esse resultado.

2.1 Informação de Fisher

A Informação de Fisher é uma função de θ , o valor esperado do quadrado da derivada com relação a θ da função de log-verossimilhança no caso uniparmétrico. Ela está diretamente relacionada aos conceitos de Eficiência e Limite inferior da variância para estimadores não viciados, conceitos que não serão abordados neste texto, mas podem ser encontrados nos textos de (BOLFARINE; SANDOVAL, 2010) e (CASELLA; BERGER, 2001). A Informação de Fisher está presente na variância assintótica de $\hat{\theta}$. Ela é definida por:

$$I(\theta) = E_{\theta_0} \left[\left(\frac{\partial}{\partial \theta} l(\theta; X) \right)^2 \right], \quad (2.1)$$

com o subscrito θ_0 para indicar que o valor do parâmetro está fixo e a esperança está sendo calculada com relação às variáveis aleatórias da amostra. Se o suporte da distribuição, o conjunto $\{x \in \mathbb{R} : f(x; \theta) > 0\}$, não depender do parâmetro θ e pudermos trocar as operações de derivação e integração duas vezes na equação acima, a Informação de Fisher também pode ser calculada por

$$E_{\theta_0}[-H(\theta)] \mid_{\theta=\theta_0} \quad (2.2)$$

em que a função $H(\theta)$ é dada pela expressão

$$H(\theta) = \frac{\partial^2}{\partial \theta^2} l(\theta; x) \quad (2.3)$$

O valor $-H(\hat{\theta})$ é chamado de Informação de Fisher Observada. Brevemente falando, no caso multiparamétrico em que temos p parâmetros, isto é, $\theta = (\theta_1, \dots, \theta_p)$, a Informação de Fisher é uma matriz cuja entrada $(i, j); i, j = 1, 2, \dots, p$ é dada por

$$[I(\theta)]_{i,j} = E_{\theta_{0i}\theta_{0j}} \left[\frac{\partial}{\partial \theta_i} l(\theta; X) \frac{\partial}{\partial \theta_j} l(\theta; X) \right] . \quad (2.4)$$

isto é, a esperança do produto das derivadas com relação aos parâmetros θ_i e θ_j respectivamente.

Exemplo 2.1 Modelo Bernoulli

Da equação (1.7), temos que a função de log-verossimilhança $l(p; x)$ é

$$l(p; x) = \sum_{i=1}^n x_i \log p + \left(n - \sum_{i=1}^n x_i \right) \log (1 - p) \quad (2.5)$$

Logo, de (2.3) segue que

$$\begin{aligned} \frac{\partial^2}{\partial p^2} l(p; x) &= -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2} \\ &= \frac{-\sum_{i=1}^n x_i (1-p)^2 - \left(n - \sum_{i=1}^n x_i \right) p^2}{p^2 (1-p)^2} \\ &= \frac{-\sum_{i=1}^n x_i (1-2p) - np^2}{p^2 (1-p)^2} \end{aligned} \quad (2.6)$$

então

$$H(p) = \frac{-\sum_{i=1}^n x_i (1-2p) - np^2}{p^2 (1-p)^2} \quad (2.7)$$

e

$$E[-H(p)] = -\frac{-np(1-2p) - np^2}{p^2 (1-p)^2} \quad (2.8)$$

Portanto, da equação (2.2) e simplificando (2.8), temos que a Informação de Fisher de p é

$$I(p) = \frac{n}{p(1-p)} \quad (2.9)$$

Exemplo 2.2 Modelo Exponencial

Da equação (1.4), temos que $l(\theta; x)$ é

$$l(\lambda; x) = n\lambda + \sum_{i=1}^n x_i \log \lambda - \sum_{i=1}^n \log (x_i!) \quad (2.10)$$

Logo, de (2.3)

$$H(\lambda) = \frac{\partial^2}{\partial \lambda^2} l(\lambda; x) = -\frac{\sum_{i=1}^n x_i}{\lambda^2} \quad (2.11)$$

O que implica que

$$E[-H(\lambda)] = -\left(-\frac{n\lambda}{\lambda^2}\right). \quad (2.12)$$

Logo, de (2.2), a Informação de Fisher de p é

$$I(\lambda) = \frac{n}{\lambda} \quad (2.13)$$

2.2 Distribuição para grandes amostras

De forma breve, enunciamos a seguir as condições de regularidade para que possamos utilizar o resultado do próximo teorema quando essas hipóteses estão satisfeitas sem dar maior atenção aos detalhes teóricos.

- 1. As variáveis aleatórias da amostra (X_1, \dots, X_n) são independentes e identicamente distribuídas, provenientes da função de densidade parametrizada por θ , $f(x; \theta)$ para todo θ no espaço paramétrico Θ .
- 2. Para diferentes valores de $\theta \in \Theta$, as funções de densidade $f(\theta; x)$ correspondem a diferentes distribuições estatísticas.
- 3. O suporte da distribuição de X_i , o conjunto $\{x_i : f(x_i; \theta) > 0\}$, não depende de θ .
- 4. O espaço paramétrico de θ é um subconjunto de \mathbb{R} e o verdadeiro valor de θ , denotado por θ_0 , não está na fronteira do espaço paramétrico.
- 5. Valem as igualdades $\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x; \theta) dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x; \theta) dx$ e $\frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(x; \theta) dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx$.
- 6. A Informação de Fisher de θ , $I(\theta)$, é finita.
- 7. A terceira derivada de $f(x; \theta)$ existe e satisfaz

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x; \theta) \right| \leq M(x) \quad (2.14)$$

em que $M(x)$ é uma função de x que não depende de θ e $E_{\theta_0}[M(x)]$ é finita em um intervalo aberto $(\theta_0 - \delta, \theta_0 + \delta)$, $\delta > 0$ centrado em θ_0 .

Note que em determinados modelos, as condições de regularidade podem estar satisfeitas para certos parâmetros e para outros não.

O próximo teorema, com adaptações, estabelece a distribuição assintótica dos estimadores de máxima verossimilhança. Novamente, o resultado será enunciado somente para o caso uniparamétrico, mas se estende para o caso multiparamétrico. A demonstração pode ser encontrada no trabalho de (MILLER, 2011).

Teorema 2.1 *Sejam θ_0 o verdadeiro valor de θ e $\hat{\theta}$ o seu estimador de máxima verossimilhança. Sob as condições de regularidade, quando o tamanho da amostra for suficientemente grande*

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, \frac{n}{I(\theta_0)}) \quad (2.15)$$

e se $g(\theta)$ é uma função inversível de θ

$$\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \sim N(0, \frac{n}{g^2(\theta)I(\theta_0)}) \quad (2.16)$$

Essas distribuições devem ser interpretadas como sendo distribuições aproximadas, e daqui em diante será omitido esse detalhe. Os tamanhos de amostra considerados grandes o suficiente para utilizar a aproximação variam e deve-se ter cuidado com cada caso observando os exemplos presentes na literatura.

Exemplo 2.3 *Modelo Poisson*

Da seção (1.3.1), segue que o E.M.V. de λ é $\hat{\lambda} = \bar{X}$. Seja λ_0 o verdadeiro valor de λ . É possível mostrar que $I(\lambda) = \frac{n}{\lambda}$. Então, pelo Teorema 2.1, temos que a distribuição de \bar{X} é

$$\sqrt{n}(\bar{X} - \lambda) \sim N(0, \lambda_0) \quad (2.17)$$

Vale notar que utilizando o Teorema Limite Central obteríamos a mesma aproximação para a distribuição de \bar{x} , em que $X \sim Poisson(\lambda_0)$.

2.3 Intervalos de Confiança e Testes de Hipóteses

Podemos construir intervalos de confiança e realizar testes de hipóteses com os E.M.V. através dos intervalos e testes de Wald, Razão de Verossimilhanças e ainda com Bootstrap. A construção de regiões de confiança, em que há mais de 1 parâmetro envolvido, será omitida.

2.3.1 Intervalos e Testes de Wald

O intervalo de confiança de Wald para θ com $(1 - \alpha) \times 100\%$ de confiança é dado por

$$\left(\hat{\theta} - z_{(1-\frac{\alpha}{2})} \sqrt{\widehat{var}(\hat{\theta})}, \hat{\theta} + z_{(1-\frac{\alpha}{2})} \sqrt{\widehat{var}(\hat{\theta})} \right) \quad (2.18)$$

em que $z_{(1-\frac{\alpha}{2})}$ é o quantil $(1 - \frac{\alpha}{2})$ da distribuição $Normal(0, 1)$ e a variância estimada de $\hat{\theta}$, denotada por $\widehat{var}(\hat{\theta})$, é dada por $-H(\hat{\theta})^{-1}$, a Informação de Fisher observada de θ calculada em $\hat{\theta}$.

O teste de Wald de nível α para testar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$ é realizado com a estatística

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\widehat{var}(\hat{\theta})}, \quad (2.19)$$

que segue a distribuição χ_1^2 . Logo, sua região de rejeição é dada por

$$[\chi_{1,(1-\alpha)}^2, +\infty) \quad (2.20)$$

em que $\chi_{1,(1-\alpha)}$ é o quantil $(1 - \alpha)$ da distribuição χ_1^2 .

Exemplo 2.4 Teste para média de uma Poisson

Suponha que tenhamos uma amostra de tamanho 20 da distribuição $Poisson(\lambda)$ e estamos interessados em estar $H_0 : \lambda = 10$ contra $H_1 : \lambda \neq 10$. Suponha que $\hat{\lambda} = \bar{x} = 15$. Da equação (2.3), temos que

$$H(\lambda) = -\frac{\sum_{i=1}^n x_i}{\lambda^2}$$

então $\widehat{var}(\hat{\lambda}) = -H(\lambda)^{-1} = \frac{\lambda^2}{\sum_{i=1}^{20} x_i} = \frac{\lambda^2}{20\bar{x}}$, pois $\sum_{i=1}^{20} \frac{x_i}{20} = \bar{x}$. Então, sob H_0 , $\widehat{var}(\hat{\lambda}) = \frac{20^2}{20 \cdot 15} = 1,33$. Logo,

$$W = \frac{(20 - 10)^2}{1,33} = 75$$

.

Considerando $\alpha = 0,05$, da equação (2.14), rejeitamos H_0 se W estiver na região $[3,84; +\infty)$. Como $W = 75$, rejeitamos a hipótese de que $\lambda = 10$.

Dessa forma, o intervalo com 95% de confiança para λ é

$$[15 - 1,96 \cdot 1,33; 15 + 1,96 \cdot 1,33] = [12,81; 17,19].$$

2.3.2 Intervalo e Teste da Razão de Verossimilhanças

A estatística da razão de verossimilhanças permite testar hipóteses da forma $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$, ou seja, as hipóteses são compostas, em que Θ_0 e Θ_1 são subconjuntos disjuntos entre si do espaço paramétrico Θ . Dessa forma, θ pode ser um vetor $(\theta_1, \dots, \theta_p)$ de p componentes.

Para calcular a estatística da razão de verossimilhanças, primeiro calculamos o E.M.V. de θ levando em conta somente os valores de θ em Θ_0 , o qual denotamos por $\hat{\theta}_0$ e depois calculamos o E.M.V. usual $\hat{\theta}$ de θ considerando todos os valores de θ no espaço paramétrico Θ . Feito isso, calculamos a estatística

$$-2 \log \lambda = -2[l(\hat{\theta}_0; x) - l(\hat{\theta}; x)] \quad (2.21)$$

Novamente sob condições de regularidade, quando o tamanho da amostra for suficientemente grande, $-2 \log \lambda \sim \chi^2_{(r-s)}$, em que r é o número de parâmetros não especificados em Θ , o espaço paramétrico de θ , e s é o número de parâmetros não especificados em Θ_0 . Dessa forma, a região de rejeição do teste de nível α é dada por $[\chi^2_{(r-s)(1-\alpha)}; +\infty)$, em que $\chi^2_{(r-s)(1-\alpha)}$ é o quantil $(1 - \alpha)$ da distribuição $\chi^2_{(r-s)}$. No caso uniparamétrico, $-2 \log \lambda \sim \chi^2_{(1)}$.

Segue imediatamente que o intervalo com $(1 - \alpha) \times 100\%$ de confiança para θ é dado pelo conjunto de valores de θ_0 tal que

$$-2[l(\theta_0; x) - l(\hat{\theta}; x)] \leq \chi^2_{1(1-\alpha)} \quad (2.22)$$

no caso uniparamétrico. Uma região de confiança, no caso multiparamétrico, pode ser calculada utilizando procedimento análogo ao que foi descrito acima.

Exemplo 2.5 *Continuação: Teste para a média de uma Poisson*

Considere o mesmo problema do exemplo anterior. Estamos interessados em testar a hipótese $H_0 : \lambda = 10$ contra $H_1 : \lambda \neq 10$. Note que $\Theta_0 = 10$, pois a hipótese nula é simples, logo, $\hat{\theta}_0 = 10$, e vimos que o E.M.V. de λ é \bar{x} . No exemplo, $\bar{x} = 15$ e suponha que $\sum_{i=1}^{20} \log(x_i!) = 545$. Da equação (1.4), temos que

$$l(\lambda; x) = -n\lambda + \log \lambda \sum_{i=1}^n x_i + \sum_{i=1}^n \log(x_i!),$$

então $l(10; x) = -83$ e $l(15; x) = -61,36$. O valor de $-2 \log \lambda$ é $-2[-83 - (-61,36)] = 43,28$. Considerando $\alpha = 0,05$, temos que a região de rejeição do teste é $[3,84; +\infty)$. Portanto, rejeitamos a hipótese de que $\lambda = 10$. Como rejeitamos a hipótese nula, construímos

um intervalo para λ com 95% de confiança dado por (2.11). O intervalo é o conjunto de valores de λ_0 tal que

$$-2[l(\lambda_0; x) - (-61, 36)] \leq 3,84$$

e com ajuda de software computacional, obtemos o intervalo $[13, 37; 16, 72]$

2.3.3 Inferência com Bootstrap

O Bootstrap é um mtodo de reamostragem no qual novas amostras são geradas a partir do conjunto de dados original. As amostras podem ser geradas selecionando aleatoriamente com reposição elementos do conjunto de dados original com igual probabilidade ou geradas através de amostras de uma distribuição em particular determinada pelos parâmetros dados pelas respectivas estimativas obtidas com o conjunto de dados original. O número de amostras geradas deve ser grande. Dessa forma, se estivermos interessados em conhecer a distribuição de um determinado estimador, aproximamos a sua distribuição através das repetições do cálculo das estimativas obtidas com as novas amostras geradas e por fim realizamos a inferência de interesse.

A metologia do Bootstrap utilizada nesse texto pode ser encontrada no trabalho de (CHERNICK, 2007), mas um texto mais aprofundado sobre Bootstrap encontra-se nos textos de (LAHIRI, 2003). O Método do Percentil, descrito abaixo em detalhes, é o método mais simples para construir intervalos de confiança e realizar testes de hipóteses aproximando a distribuição de um estimador qualquer ao redor do verdadeiro valor do parâmetro pela distribuição amostral das estimativas calculadas com as amostras geradas com as simulações de Bootstrap ao redor da estimativa da amostra original.

Método do Percentil

- Geram-se K amostras $X_k^* = (X_{1k}^*, \dots, X_{nk}^*)$, $k = 1, 2, \dots, K$ e para cada amostra X_k^* calcula-se o estimador de máxima verossimilhança $\hat{\theta}_k^*$ de θ , de modo que a distribuição de $\hat{\theta}$ em torno de θ é aproximada pela distribuição dis $\hat{\theta}_k^*$ em torno de $\hat{\theta}$.
- Para construir um intervalo com $(1 - \alpha) \times 100\%$ de confiança, encontre os quantiis $\hat{\theta}_{\frac{\alpha}{2}}^*$ e $\widehat{\theta_{(1-\frac{\alpha}{2})}^*}$ de $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_K^*$, que são respectivamente, o limite inferior e limite superior do intervalo $(\hat{\theta}_{\frac{\alpha}{2}}^*, \widehat{\theta_{(1-\frac{\alpha}{2})}^*})$, admitindo que a distribuição de $\hat{\theta}^*$ é simétrica em torno de $\hat{\theta}$.
- Para realizar um teste de hipóteses com Bootstrap, procede-se de forma análoga à construção do intervalo de confiança, mas o interesse é aproximar a distribuição de uma estatística qualquer $T(x)$. Isso é feito calculando o valor da estatística para cada amostra simulada X_k^* que denotaremos por $T(x)_k^*$ e também para os valores da

amostra original X , resultando em $K + 1$ estatísticas de teste. A região de rejeição do teste bilateral de nível α é calculado utilizando-se os quantis $T(x)_{\frac{\alpha}{2}}^*$ e $T(x)_{(1-\frac{\alpha}{2})}^*$ de $T(x)_1^*, \dots, T(x)_{K+1}^*$, de forma que a região de rejeição do teste será $(-\infty, T(x)_{\frac{\alpha}{2}}^*]$ e $[T(x)_{(1-\frac{\alpha}{2})}^*, +\infty)$.

- O p-valor é calculado contando-se quantos valores da estatística $T(x)_k^*$ são iguais ou mais extremos que o valor observado da estatística $T(x)$ na amostra original X e em seguida divide-se esse número por $K + 1$. Se o teste for bilateral, multiplica-se esse valor por 2.

Exemplo 2.6 *Continuação: Teste para a média de uma Poisson*

Considerando o problema de testar $H_0 : \lambda = 10$ contra $H_1 : \lambda \neq 10$ a partir de uma amostra da população $Poisson(\lambda)$ e construir um intervalo de confiança para λ , vamos utilizar o Método do Percentil. Os valores da amostra $X^{(2)}$ foram utilizados para esse exemplo e são apresentados na Tabela 2.

Tabela 2 – Valores da amostra $X^{(2)}$

18	16	8	12	13	17	29	10	14	13
14	20	9	12	11	14	22	19	20	9

Para realizar o Teste da Razão de Verossimilhanças foram simuladas 100 amostras a partir do conjunto de dados exposto acima e calculamos o valor de \bar{x} para cada amostra. Em seguida, calculamos o valor da estatística $-2 \log \lambda$ para cada \bar{x} das amostras simuladas com Bootstrap e para o $\bar{x} = 15$ da amostra original. Como estamos realizando um teste de nível $\alpha = 0,05$, tomamos o percentil 0,95 dos valores da estatística para construir a região de rejeição do teste, que é o valor 43,266. Portanto, a região de rejeição do teste é o intervalo $[43,266; +\infty)$. Como $-2[l(10; x) - l(15; x)] = 43,279$, rejeitamos a hipótese de que $\lambda = 10$. O intervalo de confinança para λ a 95% pode ser construído calculando o valor de \bar{x} para cada amostra e também para o valor $\bar{x} = 15$ da amostra original, e depois encontrando os quantis 0,025 e 0,975 dos $K + 1$ valores de \bar{x} , os quais foram 12,6 e 17,08 respectivamente. Logo, o intervalo com 95% de confiança para λ é $[12,6; 17,08]$.

3 Casos que não satisfazem as condições de regularidade

Vimos no capítulo anterior que certas condições de regularidade devem estar satisfeitas para podermos utilizar a distribuição Normal quando o tamanho da amostra for grande e prosseguir com a inferência acerca dos parâmetros. Quando isso não for possível, podemos usar o Bootstrap para realizar inferência. A seguir apresentamos alguns casos onde essa situação acontece.

3.1 Modelo Uniforme

No modelo $Uniforme(0, \theta)$, vimos que $X_{(n)}$ é o E.M.V. de θ . O estimador $X_{(n)}$ não tem distribuição Normal, no sentido empregado pelo Teorema 2.1, pois o suporte da distribuição $Uniforme(0; \theta)$ depende de θ . Da equação (1.11) temos que

$$0 < x < \theta, \quad \theta > 0$$

Exemplo 3.1 *Teste de hipótese para θ*

Considere a amostra $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$, para testar a hipótese $H_0 : \theta = 12$ contra $H_1 : \theta < 12$. Vamos utilizar o Teste da Razão de Verossimilhanças com o uso da distribuição Qui-quadrado e com o uso do Bootstrap. Sob H_0 , o estimador de máxima verossimilhança de θ é $\hat{\theta}_0 = 12$, pois a hipótese nula é simples. O estimador de máxima verossimilhança de θ usual é $\hat{\theta} = x_{(n)} = 10$, o máximo da amostra. Como a função de log-verossimilhança de θ é igual a $l(\theta; x) = -n \log(\theta)$ para $\theta \geq x_{(n)}$, o valor de $-2 \log \lambda = 3,65$. Caso as condições de regularidade estivessem satisfeitas, $-2 \log \lambda$ teria distribuição $\chi^2_{(1)}$ pois o número de parâmetros não especificados sob a hipótese alternativa é 1 e sob a hipótese nula é 0. Portanto, a região de rejeição do teste a nível de significância de 0,05 seria $[3,84; +\infty)$. O p-valor obtido seria 0,056. Com a utilização do Bootstrap, através de 1000 reamostragens, obtemos p-valor igual a 0,17. Nos dois casos há pouca evidência contra a hipótese nula, mas existe uma diferença notável entre os p-valores obtidos.

3.2 Modelo Semi-Circular

O modelo *Semi – Circular*(a, b) tem densidade dada por

$$f(x; a, b) = \frac{2}{\pi b} \sqrt{1 - \left(\frac{x-a}{b}\right)^2}, \quad a-b < x < a+b, \quad a \in \mathbb{R}, \quad b > 0, \quad (3.1)$$

e sua função de log-verossimilhança é dada por:

$$l(a, b; x) = \frac{1}{2} \sum_{i=1}^n \log(b^2 - (x_i - a)^2) - 2n \log b + n \log \frac{2}{\pi} \quad (3.2)$$

Logo, a derivada parcial de $l(a, b; x)$ com relação ao parâmetro a é

$$\frac{\partial}{\partial a} l(a, b; x) = \sum_{i=1}^n \frac{x_i - a}{b^2 - (x_i - a)^2}, \quad (3.3)$$

No entanto,

$$I(a) = E_a \left[\left(\frac{\partial}{\partial a} l(a, b; x) \right)^2 \right] = \infty \quad (3.4)$$

Isto é, a Informação de Fisher de a não é finita. Analogamente, isso ocorre para o parâmetro b . Dessa forma, os estimadores de máxima verossimilhança de a e b , \hat{a} e \hat{b} respectivamente, não têm distribuição Normal, pois tanto $I(a)$ quanto $I(b)$ não são finitas.

Exemplo 3.2 Teste de hipótese para a

Considere novamente a amostra $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$, mas agora de $X \sim \text{Semi} - \text{Circular}(a, 5)$ para testar a hipótese $H_0 : a = 5, 9$ contra a hipótese $H_1 : a \neq 5, 9$. Sob a hipótese nula, $\hat{a}_0 = 5, 9$ e $\hat{a} = 5, 5$ é o E.M.V. usual de a . De (3.2) temos que o valor de $-2 \log \lambda = 1, 16$. Caso as condições de regularidade estivessem satisfeitas, $-2 \log \lambda \sim \chi^2_{(1)}$ e o p-valor obtido seria 0,28. Com o uso do Bootstrap, realizando 1000 reamostragens sob a hipótese nula, obtemos p-valor de 0,169. Nos dois casos não há evidência contra H_0 .

3.3 Modelos de Mistura

Em um modelo de mistura pressupomos a existência de g grupos, G_1, \dots, G_g , de modo que $X_j, j = 1, 2, \dots, n$ tem probabilidade igual a λ_i de pertencer ao grupo G_i e $\sum_{i=1}^g \lambda_i = 1$. Veja (MCLACHLAN; BASFORD, 1987). X_j pode ser um vetor aleatório de tamanho $p \geq 1$, mas neste texto vamos nos ater ao caso univariado. Se X_j pertencer ao grupo G_i , então sua densidade será dada por $f_i(x_j)$. Tome $Z_j = (z_{1j}, \dots, z_{gj}) \sim \text{Multinomial}(1; \lambda_1, \dots, \lambda_g)$

e se $z_{ij} = 1$ então a densidade de X_j será $f_i(x_j)$, $\sum_{i=1}^g z_{ij} = 1$ e, portanto, X_j pertence a um único grupo. A densidade conjunta de (X_j, Z_j) é

$$\begin{cases} \lambda_1 f_1(x_j) & \text{se } z_{1j} = 1 \\ \vdots & \vdots \\ \lambda_g f_g(x_j) & \text{se } z_{gj} = 1 \end{cases} \quad (3.5)$$

No entanto, se na amostra não observamos os valores de Z_j , ou seja, se não soubermos a verdadeira procedência de cada X_j , os dados são incompletos e trabalhamos apenas com os valores de X_j . Logo, estamos interessados em estimar as proporções λ_i . A densidade marginal de X_j é dada por:

$$g(x_j, \lambda) = \sum_{i=1}^g \lambda_i f_i(x_j), \quad \lambda = (\lambda_1, \dots, \lambda_g) \quad (3.6)$$

Em um modelo de mistura composto de g grupos, $g > 1$, se existe $\lambda_i = 0$ para algum i , então o verdadeiro valor de λ_i está na fronteira do espaço paramétrico. Portanto, as condições de regularidade não estão satisfeitas para λ_i .

Exemplo 3.3 Modelo Exponencial

Para um modelo de mistura com $g = 2$ componentes e $f_i(x) = \frac{1}{\mu_i} e^{-\frac{x}{\mu_i}}$, $i = 1, 2$, isto é, se X_j pertencer ao grupo i então $X_j \sim \text{Exponencial}(\mu_i)$, e $g(x_j)$ é dado por

$$g(x_j, \lambda) = \lambda_1 \frac{1}{\mu_1} e^{-\frac{x_j}{\mu_1}} + \lambda_2 \frac{1}{\mu_2} e^{-\frac{x_j}{\mu_2}} \quad (3.7)$$

3.3.1 Função de Verossimilhança com dados completos

O E.M.V. do vetor de parâmetros $\psi = (\lambda, \theta)$ do modelo de mistura, em que $\theta = (\theta_1, \dots, \theta_g)$ e cada θ_i é o vetor de parâmetros da densidade $f_i(x)$, são calculados considerando a função de log-verossimilhança com os dados completos (X_j, Z_j) , mas como não observamos os valores de Z_j , estes são estimados com o Algoritmo E.M. juntamente com o E.M.V. de ψ . Para um texto mais detalhado sobre o Algoritmo E.M. consulte o trabalho de (MCLACHLAN; KRISHNAN, 2008).

Denotamos a função de log-verossimilhança com os dados completos por $l_C(\psi)$, dada pela expressão:

$$l_C(\psi) = \sum_{i=1}^g \sum_{j=1}^n z_{ij} [\log \lambda_i + \log f_i(x_j)] \quad (3.8)$$

Enfatizamos o fato de que $\sum_{i=1}^g \lambda_i = 1$ e portanto estimamos apenas $g - 1$ dos $\lambda_1, \dots, \lambda_g$, sendo que um deles, digamos λ_g é dado por $1 - \sum_{i=1}^{g-1} \lambda_i$.

3.3.2 Estimação de ψ com o Algoritmo E.M.

O Algoritmo E.M. consiste de duas etapas E-STEP e M-STEP. O procedimento seguinte está sendo considerado somente para o cálculo dos E.M.V. dos parâmetros de um modelo de mistura descrito por (3.6). Na iteração $(k+1)$ a etapa E-STEP calcula o valor $b_{ij}^{*(k+1)}$, que é a estimativa de z_{ij} na iteração $(k+1)$, e é também o valor esperado de z_{ij} condicionado ao valor de ψ e x_j , dado por:

$$b_{ij}^{*(k+1)} = \frac{\lambda_i^{(k)} f_i(x_j)}{\sum_{t=1}^g \lambda_t^{(k)} f_t(x_j)}. \quad (3.9)$$

São utilizados os valores $\lambda_i^{(k)}$ e $\theta_i^{(k)}$ que foram obtidos na iteração anterior (k) para calcular $b_{ij}^{*(k+1)}$. Note que o valor de $f_i(x_j)$ requer os valores do vetor θ_i para ser calculado. Na iteração (1) escolhemos arbitrariamente os valores de $\lambda_i^{(0)}$ e $\theta_i^{(0)}$.

Para prosseguir com o procedimento iterativo a etapa M-STEP calcula o estimador de máxima verossimilhança de ψ na iteração $(k+1)$, que é qualquer valor $\psi^{(k+1)}$ que maximiza a função

$$l_C(\psi)^{(k+1)} = \sum_{i=1}^g \sum_{j=1}^n b_{ij}^{*(k+1)} [\log \lambda_i + \log f_i(x_j)] \quad (3.10)$$

lembrando que $\psi^{(k+1)} = (\lambda^{(k+1)}, \theta^{(k+1)})$. O valor $\lambda_i^{(k+1)}$ é dado por

$$\lambda_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k+1)}}{n}, i = 1, 2, \dots, g-1 \quad (3.11)$$

e $\lambda_g^{(k+1)} = 1 - \sum_{i=1}^{g-1} \lambda_i^{(k+1)}$. Em alguns casos é possível encontrar uma apresentação analítica para $\theta_i^{(k)}$. Considerando o exemplo anterior do modelo de mistura Exponencial, em que $g = 2$ e $\psi = (\lambda_1, \lambda_2, \mu_1, \mu_2)$, temos que

$$\mu_1^{(k+1)} = \frac{\sum_{i=1}^n b_{1j}^{*(k+1)} x_j}{\sum_{i=1}^n b_{1j}^{*(k+1)}} \quad (3.12)$$

e

$$\mu_2^{(k+1)} = \frac{\sum_{i=1}^n (1 - b_{1j}^{*(k+1)}) x_j}{\sum_{i=1}^n (1 - b_{1j}^{*(k+1)})} \quad (3.13)$$

Exemplo 3.4 Continuação: Modelo de Mistura Exponencial

Considere a amostra $X^{(3)}$, os valores são apresentados na Tabela 3.

Suponha que $X^{(3)}$ seja proveniente de um modelo de mistura dado por (3.7). Então temos $g = 2$ grupos e as componentes são Exponenciais. O vetor de parâmetros no nosso

Tabela 3 – Valores da amostra $X^{(3)}$

18.63	0.95	10.72	1.27	16.61	6.9	7.3	0.41
5.9	1.92	31.41	1.1	11.35	10.61	19.06	

exemplo é $\psi = (\lambda_1, \lambda_2, \mu_1, \mu_2)$, mas utilizaremos $\psi = (\lambda_1, \mu_1, \mu_2)$ pois $\lambda_2 = 1 - \lambda_1$. Para iniciar o Algoritmo E.M. vamos utilizar $\lambda_1 = 0.5$, $\mu_1 = q_{(1)} = 1.6$ e $\mu_2 = q_{(3)} = 13.98$, em que $q_{(1)}$ e $q_{(3)}$ são, respectivamente, o primeiro e terceiro quartis da amostra. Para ilustrar o procedimento, a seguir mostramos a realização das 2 primeiras iterações.

- **Iteração 1:** Na iteração 1 realizamos primeiramente a etapa E-STEP calculando os valores dos $b_{ij}^{*(1)}$, $i = 1, 2; j = 1, \dots, 15$. Da equação (3.9) segue que valor de $b_{1,1}^{*(1)}$ é calculado da seguinte forma:

$$b_{1,1}^{*(1)} = \frac{\lambda_1^{(0)} f_1(x_1)}{\sum_{i=1}^2 \lambda_i^{(0)} f_i(x_1)} = \frac{0.5 \frac{1}{1.6} e^{-\frac{18.63}{1.6}}}{0.5 \frac{1}{1.6} e^{-\frac{18.63}{1.6}} + 0.5 \frac{1}{13.98} e^{-\frac{18.63}{13.98}}} = 0,000291$$

Os demais valores de $b_{1j}^{*(1)}$ são calculados analogamente e foram arredondados com 6 casas decimais para facilitar a demonstração do cálculo. Esses valores estão na Tabela 4. Em seguida, realizamos o passo M-STEP calculando os valores de $\lambda_1^{(1)}$, $\mu_1^{(1)}$ e $\mu_2^{(1)}$. O cálculo é feito da seguinte forma:

De (3.11), segue que

$$\lambda_1^{(1)} = \frac{\sum_{j=1}^n b_{1,j}^{*(1)}}{n} = \frac{0,000291 + 0,837784 + \dots + 0,000229}{15} = 0,314018$$

Das equações (3.12) e (3.13) segue que

$$\mu_1^{(1)} = \frac{\sum_{j=1}^n b_{1j}^{*(1)} x_j}{\sum_{j=1}^n b_{1j}^{*(1)}} = \frac{0,000291 \cdot 18,63 + \dots + 0,000229 \cdot 19,06}{0,000291 + \dots + 0,000229} = 1,868326$$

e

$$\mu_2^{(1)} = \frac{\sum_{i=1}^n (1 - b_{1j}^{*(1)}) x_j}{\sum_{i=1}^n (1 - b_{1j}^{*(1)})} = \frac{0,999709 \cdot 18,63 + \dots + 0,999771 \cdot 19,06}{0,999709 + \dots + 0,999771} = 13,152893$$

- **Iteração 2:** Na iteração 2 utilizamos os valores de $\lambda_1^{(1)}$, $\mu_1^{(1)}$ e $\mu_2^{(1)}$ para obter os valores $b_{ij}^{*(2)}$ na etapa E-STEP. Dessa forma, o valor de $b_{1,1}^{*(2)}$ é dado por:

$$b_{1,1}^{*(2)} = \frac{\lambda_1^{(1)} f_1(x_1)}{\sum_{i=1}^2 \lambda_i^{(1)} f_i(x_1)} = \frac{0,314018 \cdot \frac{1}{1,868326} e^{-\frac{18,63}{1,868326}}}{0,314018 \cdot \frac{1}{1,868326} e^{-\frac{18,63}{1,868326}} + 0,685982 \cdot \frac{1}{1,868326} e^{-\frac{18,63}{1,868326}}} = 0,00062$$

Da mesma forma procedemos para realizar o cálculo dos demais $b_{1j}^{*(2)}$. Os seus valores são mostrados na Tabela 4. Em seguida, na etapa M-STEP calculamos os valores $\lambda_1^{(2)}$, $\mu_1^{(2)}$ e $\mu_2^{(2)}$ utilizando os valores $b_{ij}^{*(2)}$. Portanto, temos que

$$\lambda_1^{(2)} = \frac{\sum_{j=1}^n b_{1,j}^{*(2)}}{n} = \frac{0,000620 + 0,675671 + \dots + 0,000509}{15} = 0,249483$$

$$\mu_1^{(2)} = \frac{\sum_{j=1}^n b_{1,j}^{*(2)} x_j}{\sum_{j=1}^n b_{1,j}^{*(2)}} = \frac{0,000620 \cdot 18,63 + \dots + 0,000509 \cdot 19,06}{0,000620 + \dots + 0,000509} = 1.852055$$

e

$$\mu_2^{(2)} = \frac{\sum_{i=1}^n (1 - b_{1j}^{*(2)}) x_j}{\sum_{i=1}^n (1 - b_{1j}^{*(2)})} = \frac{0,999380 \cdot 18,63 + \dots + 0,999491 \cdot 19,06}{0,999380 + \dots + 0,999491} = 12.187966$$

Tabela 4 – Valores dos $b_{1j}^{*(1)}$ e $b_{1j}^{*(2)}$ para $j = 1, \dots, 15$.

Observação	$b_{1j}^{*(1)}$	$b_{1j}^{*(2)}$
1	0,000291	0,000620
2	0,837784	0,675671
3	0,022631	0,022921
4	0,812253	0,642677
5	0,000888	0,001567
6	0,160931	0,119376
7	0,133229	0,101376
8	0,874429	0,727490
9	0,250144	0,176661
10	0,751189	0,571633
11	0,000000	0,000002
12	0,826182	0,660398
13	0,016076	0,017262
14	0,024017	0,024080
15	0,000229	0,000509

A Tabela 5 expõe os valores de λ_1 , μ_1 e μ_2 nas primeiras 15 iterações. Na iteração 0 são mostrados os valores iniciais de ψ . $\hat{\psi}$ converge para o vetor (0.12, 1.05, 10.74) na 15ª iteração. O valor da log-verossimilhança em $\hat{\psi}$ é -48.71 .

Geralmente a função de log-verossimilhança de modelos de mistura apresentam múltiplos pontos de máximo e consequentemente múltiplos valores para o E.M.V. de ψ . É necessário saber qual das estimativas proporciona o maior valor para função de log-verossimilhança. Também requerem uma grande quantidade de observações para obter boas estimativas.

A estimativa de λ_1 , no nosso exemplo, foi igual a 0.12. Se quisermos testar a hipótese de que $H_0 : \lambda_1 = 0$ contra $H_1 : \lambda_1 \neq 0$, não poderemos utilizar o teste de Wald e também não poderemos usar o Teste da Razão de Verossimilhanças com a distribuição Qui-quadrado, pois 0 está na fronteira do espaço paramétrico de λ_1 que é o intervalo $[0, 1]$. O mesmo vale para encontrar um intervalo de confiança para λ_1 . Em casos especiais é possível obter a distribuição assintótica dos estimadores de λ_i e realizar testes de hipóteses a respeito do número de grupos. No entanto, esse texto não dará mais atenção a esses casos, pois todos serão tratados de modo geral. Para mais detalhes, consulte o trabalho de (MCLACHLAN; BASFORD, 1987).

Tabela 5 – Valores de $\psi^{(k)}$ em cada iteração $k = 0, 1, \dots, 15$

Iteração	λ_1	μ_1	μ_2
0	0.5	1.6	13.98
1	0,314018	1,868326	13,15289
2	0,249483	1,852055	12,18797
3	0,210477	1,723479	11,71161
4	0,184068	1,544929	11,42861
5	0,165560	1,367590	11,24456
6	0,152493	1,230826	11,11688
7	0,143093	1,146456	11,02253
8	0,136129	1,101722	10,94997
9	0,130871	1,079136	10,89379
10	0,126870	1,067315	10,85053
11	0,123806	1,060565	10,81727
12	0,121447	1,056320	10,79166
13	0,119619	1,053414	10,77184
14	0,118196	1,051307	10,75644
15	0,117082	1,049734	10,74440

Exemplo 3.5 *Teste de hipótese para g*

Considere a amostra cujos valores são apresentados na Tabela 3. Estamos interessados em testar a hipótese $H_0 : g = 1$ contra a hipótese $H_1 : g = 2$. Vamos utilizar o Teste da Razão de Verossimilhanças com a distribuição Qui-Quadrado e também com o Bootstrap. Se $g = 1$, então a distribuição de cada observação da amostra é *Exponencial*(μ) e portanto $\hat{\mu} = \bar{x} = 9,61$ é o E.M.V. de μ . Para $g = 2$ obtemos $\hat{\psi} = (0,12; 1,05; 10,74)$. O valor de $-2 \log \lambda = 0,45$. Para calcular este valor, utilizamos a função de log-verossimilhança da função Exponencial dada por (2.10) e da log-verossimilhança dada por (1.2) com a densidade do modelo de mistura Exponencial com 2 grupos. Com a distribuição $\chi^2_{(1)}$ obtemos p-valor igual a 0,5 e utilizando o Bootstrap obtemos p-valor igual a 0,25 através de 1000 reamostragens sob a hipótese nula. Portanto, não há evidência contra H_0 nos dois casos, apesar do resultado obtido com a distribuição Qui-quadrado não ser válida. Constatamos que há uma discrepância bem evidente entre os p-valores obtidos.

3.3.3 *Mistura de Normais*

No contexto de modelos de mistura, talvez o caso mais importante seja o caso em que há componentes da distribuição Normal. Se $X_j, j = 1, 2, \dots, n$ pertence ao grupo $G_i, i = 1, 2, \dots, g$ então $X_j \sim \text{Normal}(\mu_i, \sigma_i^2)$. Logo, da equação (3.6), temos que $g(x_j, \lambda)$ é da forma

$$g(x_j, \lambda) = \sum_{i=1}^g \frac{\lambda_i}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_j - \mu_i)^2}{2\sigma_i^2}} \quad (3.14)$$

O vetor de parâmetros é $\psi = (\lambda, \mu, \sigma)$ em que $\mu = (\mu_1, \dots, \mu_g)$ e $\sigma = (\sigma_1, \dots, \sigma_g)$. Seguindo a mesma direção do que foi feito na subseção anterior para obter as estimativas de θ_i com o algoritmo E.M., será necessário apenas determinar os valores de μ_i e σ_i , pois as estimativas de λ_i e z_{ij} são as mesmas para todos os casos no Algoritmo E.M.. Dessa forma, na etapa M-STEP, temos que as estimativas de máxima verossimilhança de μ_i e σ_i na iteração $(k+1)$ são, respectivamente, dadas por:

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k)} x_j}{\sum_{j=1}^n b_{ij}^{*(k)}} \quad (3.15)$$

e

$$\sigma_i^{2(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k)} (x_j - \mu_i^{(k)})^2}{\sum_{j=1}^n b_{ij}^{*(k)}} \quad (3.16)$$

Se supormos que $\sigma_i^2 = \sigma^2$ para todo $i = 1, 2, \dots, g$, será necessário apenas determinar o valor de σ^2 , então a estimativa de máxima verossimilhança de σ^2 na iteração $(k + 1)$ se reduz a:

$$\sigma^{2(k+1)} = \frac{\sum_{i=1}^g \sum_{j=1}^n b_{ij}^{*(k)} (x_j - \mu_i^{(k)})^2}{n} \quad (3.17)$$

Para mais detalhes consulte (MCLACHLAN; BASFORD, 1987).

3.3.4 Mistura de distribuições Poisson

Considere o modelo de mistura dado por (3.6), se X_j pertence ao grupo $G_i, i = 1, 2, \dots, g$ então a função de probabilidade de X_j é dada por:

$$f_i(x_j) = \frac{e^{-\mu_i} \mu_i^{x_j}}{x_j!}. \quad (3.18)$$

Logo, $g(x_j, \lambda)$ é dada pela expressão:

$$g(x_j, \lambda) = \sum_{i=1}^g \lambda_i \frac{e^{-\mu_i} \mu_i^{x_j}}{x_j!} \quad (3.19)$$

Assim, na etapa M-STEP, a estimativa de máxima verossimilhança de μ_i na iteração $(k + 1)$ também é dada por:

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k+1)} x_j}{\sum_{j=1}^n b_{ij}^{*(k+1)}} \quad (3.20)$$

Novamente, este valor deve ser calculado sucessivamente para cada iteração $(k + 1)$ na etapa M-STEP.

Exemplo 3.6 *Mistura de duas distribuições Poisson*

Considerando um modelo de mistura com $g = 2$ grupos dado por (3.19) e os valores da amostra $X^{(4)} = (1, 2, 3, 4, 5, 6)$, vamos realizar as 2 primeiras iterações do Algoritmo E.M. e determinar os valores de λ_1 , μ_1 e μ_2 nessas iterações. Da equação (3.19), temos que

$$g(x_j; \lambda) = \lambda_1 \frac{e^{\mu_1} \mu_1^{-x_j}}{x_j!} + \lambda_2 \frac{e^{\mu_2} \mu_2^{-x_j}}{x_j!} \quad (3.21)$$

de modo que $\lambda_2 = 1 - \lambda_1$.

Os valores utilizados para iniciar o algoritmo E.M. são $\lambda_1^{(0)} = 0,5$, $\mu_1^{(0)} = 2,25$, o primeiro quartil dos valores amostra, e $\mu_2^{(0)} = 4,75$, o terceiro quartil dos valores da amostra. Lembrando que não é necessário determinar o valor de λ_2 pois este é dado por $1 - \lambda_1$.

- **Iteração 1:** Na etapa E-STEP calculamos os valores dos $b_{ij}^{*(1)}$, $i = 1, 2; j = 1, \dots, 6$. O cálculo de $b_{1,3}^{*(1)}$ é da seguinte forma:

De (3.9)

$$b_{1,3}^{*(1)} = \frac{\lambda_1^{(0)} f_1(x_3)}{\sum_{i=1}^2 \lambda_i^{(0)} f_i(x_3)} = \frac{0,5 \cdot \frac{e^{-2,25} 2,25^3}{3!}}{0,5 \cdot \frac{e^{-2,25} 2,25^3}{3!} + 0,5 \cdot \frac{e^{-4,75} 4,75^3}{3!}} = 0,625579$$

e

$$b_{2,3}^{*(1)} = 1 - b_{1,3}^{*(1)} = 0,374421$$

Os demais valores de $b_{ij}^{*(1)}$ são calculados de forma análoga e os valores são apresentados na Tabela 6. Em seguida, na etapa M-STEP, determinamos os valores de $\lambda_1^{(1)}$, $\mu_1^{(1)}$ e $\mu_2^{(1)}$. Das equações (3.11) e (3.20) segue que:

$$\lambda_1^{(1)} = \frac{\sum_{j=1}^n b_{1,j}^{*(1)}}{n} = \frac{0,516919 + \dots + 0,447971}{6} = 0,482814$$

$$\mu_1^{(1)} = \frac{\sum_{j=1}^n b_{1,j}^{*(1)} x_j}{\sum_{j=1}^n b_{1,j}^{*(1)}} = \frac{0,625579 \cdot 1 + \dots + 0,341573 \cdot 6}{0,625579 + \dots + 0,341573} = 3,155024$$

e

$$\mu_2^{(1)} = \frac{\sum_{j=1}^n b_{2,j}^{*(1)} x_j}{\sum_{j=1}^n b_{2,j}^{*(1)}} = \frac{0,374421 \cdot 1 + \dots + 0,658427 \cdot 6}{0,374421 + \dots + 0,658427} = 3,822048$$

lembrando que $b_{2j}^{*(1)} = 1 - b_{1j}^{*(1)}$.

- **Iteração 2:** Na etapa E-STEP calculamos os valores dos $b_{ij}^{*(2)}$, $i = 1, 2; j = 1, \dots, 6$ utilizando os valores de $\lambda_1^{(1)}$, $\mu_1^{(1)}$ e $\mu_2^{(1)}$. Logo, calculamos $b_{1,3}^{*(2)}$ da seguinte forma:

$$b_{1,3}^{*(2)} = \frac{\lambda_1^{(1)} f_1(x_1)}{\sum_{i=1}^2 \lambda_i^{(1)} f_i(x_1)} = \frac{0,482814 \cdot \frac{e^{-3,155024} 3,155024^3}{3!}}{0,482814 \cdot \frac{e^{-3,155024} 3,155024^3}{3!} + 0,517186 \cdot \frac{e^{-3,822048} 3,822048^3}{3!}} = 0,48927$$

e

$$b_{2,3}^{*(2)} = 1 - b_{1,3}^{*(2)} = 0,51073$$

Os demais valores de $b_{ij}^{*(2)}$ são calculados de forma análoga e são apresentados na Tabela 6. Em seguida, na etapa M-STEP, determinamos os valores de $\lambda_1^{(2)}$, $\mu_1^{(2)}$ e $\mu_2^{(2)}$. Da mesma forma que na iteração anterior, temos que

$$\lambda_1^{(2)} = \frac{\sum_{j=1}^n b_{1,j}^{*(2)}}{n} = \frac{0,625579 + \dots + 0,341573}{6} = 0,4824$$

$$\mu_1^{(2)} = \frac{\sum_{j=1}^n b_{1j}^{*(2)} x_j}{\sum_{j=1}^n b_{1j}^{*(2)}} = \frac{0,516919 \cdot 1 + \dots + 0,447971 \cdot 6}{0,516919 + \dots + 0,447971} = 3,4166$$

e

$$\mu_2^{(2)} = \frac{\sum_{j=1}^n b_{2j}^{*(2)} x_j}{\sum_{j=1}^n b_{2j}^{*(2)}} = \frac{0,483081 \cdot 1 + \dots + 0,552029 \cdot 6}{0,483081 + \dots + 0,552029} = 3,577728$$

Lembrando que $b_{2j}^{*(2)} = 1 - b_{1j}^{*(2)}$.

Tabela 6 – Valores dos $b_{1j}^{*(1)}$ e $b_{1j}^{*(2)}$ para $j = 1, \dots, 6$.

Observação	$b_{1j}^{*(1)}$	$b_{1j}^{*(2)}$
1	0,625579	0,516919
2	0,569394	0,503097
3	0,511363	0,489270
4	0,453024	0,475459
5	0,395948	0,461686
6	0,341573	0,447971

4 Aplicações

Neste capítulo encontram-se algumas aplicações de casos em que as condições de regularidade não estão satisfeitas para determinados parâmetros e que, conseqüentemente, os correspondentes estimadores de máxima verossimilhança não têm distribuição Normal, no mesmo sentido em que foi empregado no capítulo anterior.

4.1 Modelos de Mistura

A seguir, apresentamos duas aplicações de modelos de mistura. Considerando que o modelo de mistura com densidade dada por (3.6) em que há g grupos é adequado aos dados em estudo, estamos interessados em saber qual é a quantidade mínima de grupos que compõem o modelo de mistura. Enfatizamos o fato de que o modelo de mistura não satisfaz as condições de regularidade para λ_i , no entanto, isso não é necessariamente verdadeiro para os outros parâmetros θ_i da densidade $f_i(x)$ correspondente ao grupo G_i . Portanto, com relação aos parâmetros λ_i , não podemos conduzir a inferência da teoria padrão com a aproximação pela distribuição Normal, como foi dito anteriormente.

A primeira aplicação é um problema de análise de clusters em que estamos interessados em saber se o número de mortes causadas por Tuberculose, relativo ao número de habitantes no Brasil no ano de 2015 é o mesmo para todas as Unidades da Federação ou se existem grupos, clusters, de Unidades da Federação que diferem no número de mortes causadas por Tuberculose. A segunda aplicação é um problema de identificação de outliers, ou observações discrepantes, em um problema de regressão linear. O objetivo é utilizar o modelo de mistura para testar a hipótese de que não há o grupo do qual provém os outliers contra a hipótese de que há esse grupo, no conjunto de dados, através da avaliação do número de grupos que compõem o modelo. De forma que se não rejeitarmos a hipótese de que há o grupo do qual provém os outliers, então será admitido que os dados são provenientes do grupo do modelo de regressão, ou seja, todos as observações do conjunto de dados se ajustam ao modelo de regressão.

4.1.1 Classificação do número de mortes por Tuberculose

Seguindo a abordagem de (SYMONS; GRIMSON; YUAN, 1983) para classificar eventos raros, de forma análoga estamos fazendo a classificação do número de mortes causadas por tuberculose nas Unidades Federativas do Brasil no ano de 2015. Admite-se que o número de mortes causadas por Tuberculose em cada Unidade da Federação do Brasil, denotado por $x_j, j = 1, 2, \dots, 27$, no ano de 2015, segue a distribuição $Poisson(\mu_j')$, em

que $\mu_j' = \mu_i P_j, i = 1, \dots, g$ depende do número de habitantes P_j da Unidade da Federação j e da probabilidade de ocorrer morte causada por Tuberculose μ_i referente ao grupo G_i a que essa Unidade da Federação pertence, segundo o número de mortes causadas por Tuberculose. Será admitido que, se houver mais de um grupo, o grupo cujo valor μ_i seja o menor de todos será considerado o grupo com número baixo de mortes causadas por tuberculose. O grupo cujo valor μ_i seja o maior de todos será considerado o grupo com número alto de mortes causadas por tuberculose. Por último, os grupos com o valor μ_i que eventualmente ficarem entre o menor e o maior valores serão considerados com número intermediário de mortes causadas por Tuberculose. O número de habitantes para cada Unidade da Federação tomado é uma estimativa da população residente para o ano de 2015 segundo a RIPSa (Rede Interagencial de Informações para Saúde), mas no exemplo será considerado como o verdadeiro número de habitantes. A Tabela 7 apresenta o número de mortes causadas por Tuberculose e a estimativa da população residente de cada Unidade da Federação no ano de 2015.

Tabela 7 – Número de mortes causadas por tuberculose e estimativa da população residente segundo a Unidade da Federação. Fonte: www.datasus.gov.br

Unid. da Federação	Mortes (x_j)	População (P_j)	Freq. rel. de mortes
Rondônia	27	1768162	0,001527%
Acre	20	803503	0,002489%
Amazonas	125	3938325	0,003174%
Roraima	4	505674	0,000791%
Pará	208	8175156	0,002544%
Amapá	14	766692	0,001826%
Tocantins	9	1515134	0,000594%
Maranhão	150	6904298	0,002173%
Piauí	44	3203248	0,001374%
Ceará	208	8905190	0,002336%
Rio Grande do Norte	65	3442158	0,001888%
Paraíba	88	3972175	0,002215%
Pernambuco	418	9345638	0,004473%
Alagoas	76	3340528	0,002275%
Sergipe	46	2242948	0,002051%
Bahia	411	15203851	0,002703%
Minas Gerais	237	20869033	0,001136%
Espírito Santo	77	3929925	0,001959%
Rio de Janeiro	832	16550009	0,005027%
São Paulo	872	44396460	0,001964%
Paraná	126	11163023	0,001129%
Santa Catarina	50	6819124	0,000733%
Rio Grande do Sul	294	11247923	0,002614%
Mato Grosso do Sul	49	2651222	0,001848%
Mato Grosso	71	3265468	0,002174%
Goiás	63	6610683	0,000953%
Distrito Federal	26	2914830	0,000892%

No problema em questão, segue que da equação (3.6) o modelo de mistura tem densidade dada por:

$$g(x_j, \lambda) = \sum_{i=1}^g \lambda_i \frac{e^{-\mu_i P_j} (\mu_i P_j)^{x_j}}{x_j!} \quad (4.1)$$

Desta forma, estamos interessados em determinar o número mínimo de grupos que compõem o modelo de mistura e para isso precisamos estimar os valores dos λ_i e μ_i , $i = 1, 2, \dots, g$. Considerando a densidade dada por (4.1), a função de log-verossimilhança completa de $\psi = (\lambda, \mu)$ é dada por:

$$l_C(\psi) = \sum_{i=1}^g \sum_{j=1}^n z_{ij} [\log \lambda_i - \mu_i P_j + x_j \log(\mu_i P_j) - \log(x_j!)] \quad (4.2)$$

em que $\mu = (\mu_1, \dots, \mu_g)$ e cada um dos x_j e P_j estão fixos. Em seguida, obtemos as equações das estimativas de máxima verossimilhança dos μ_i na iteração $(k + 1)$ maximizando $l_C(\psi)$ com relação a μ_i e substituindo os z_{ij} pelos $b_{ij}^{*(k+1)}$. Ressaltamos que a função de log-verossimilhança completa é utilizada somente no algoritmo E.M. para determinar os valores das estimativas. Já para realizar o teste da Razão de Verossimilhanças utilizamos a função de log-verossimilhança usual dada por (1.2). De (4.1) temos que

$$l(\lambda_1, \mu_1, \mu_2; x) = \sum_{j=1}^n \log \left(\sum_{i=1}^g \lambda_i \frac{e^{-\mu_i P_j} (\mu_i P_j)^{x_j}}{x_j!} \right) \quad (4.3)$$

Utilizamos o algoritmo E.M. para calcular o valor dos E.M.V. dos λ_i e μ_i . As estimativas de λ_i são as usuais, dadas por (3.11), e somente para μ_i há uma nova expressão. Portanto, na iteração $(k + 1)$, utilizamos as estimativas:

$$b_{ij}^{*(k+1)} = \frac{\lambda_i^{(k)} f_i(x_j)}{\sum_{t=1}^g \lambda_t^{(k)} f_t(x_j)} \quad (4.4)$$

e

$$\lambda_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k+1)}}{n}, i = 1, 2, \dots, g - 1 \quad (4.5)$$

como visto anteriormente. Lembrando que estimamos apenas $g - 1$ dos λ_i pois o último é obtido por $1 - \sum_{i=1}^{g-1} \lambda_i$. As estimativas de μ_i na iteração $(k + 1)$ são dadas por:

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k+1)} x_j}{\sum_{j=1}^n b_{ij}^{*(k+1)} P_j} \quad (4.6)$$

No caso em que $g = 1$, ou seja, supondo a existência de apenas 1 grupo, $\hat{\mu}_1 = \frac{\sum_{j=1}^n x_j}{\sum_{j=1}^n P_j}$ é o E.M.V. de μ_1 . Logo, não é necessário utilizar o algoritmo E.M. pois encontramos o E.M.V. de μ_1 com apresentação analítica da forma usual e não temos que estimar λ_1 pois este é igual a 1.

São apresentados na Tabela 8 os valores de $\lambda_1^{(k)}$, $\mu_1^{(k)}$ e $\mu_2^{(k)}$ nas primeiras 3 iterações. Os valores utilizados para iniciar o algoritmo E.M. foram $\lambda_1^{(0)} = 0.2$, $\mu_1^{(0)} = 1,003667 \times 10^{-5}$ e $\mu_2^{(0)} = 2,254826 \times 10^{-5}$.

Tabela 8 – Valores das estimativas nas 3 primeiras iterações.

Iteração	$\lambda_1^{(k)}$	$\mu_1^{(k)}$	$\mu_2^{(k)}$
0	0,2000000	$1,003667 \times 10^{-5}$	$2,254826 \times 10^{-5}$
1	0,2996757	$1,046809 \times 10^{-5}$	$2,686788 \times 10^{-5}$
2	0,3382299	$1,060741 \times 10^{-5}$	$2,699490 \times 10^{-5}$
3	0,3417802	$1,062478 \times 10^{-5}$	$2,700376 \times 10^{-5}$

Começando a análise considerando um modelo simples para os dados, ou seja, há somente $g = 1$ grupo, obtemos $\hat{\mu}_1 = 2,25483 \times 10^{-5}$ e da equação (4.3), temos que o valor da função de log-verossimilhança neste ponto é $-631,2993$, em que $g = 1$ e $n = 27$. Se considerarmos a existência de $g = 2$ grupos, obtemos $\hat{\lambda}_1 = 0,3422$, $\hat{\mu}_1 = 1,06271 \times 10^{-5}$ e $\hat{\mu}_2 = 2,70048 \times 10^{-5}$. O valor da função de log-verossimilhança neste caso é $-367,4788$. Na equação (4.3) temos $g = 2$ e $n = 27$. O valor da estatística de $-2 \log \lambda$ é $527,461$. O teste foi realizado com o Bootstrap utilizando o Método do Percentil, descrito na subseção 2.3.3. Foram geradas 350 amostras para calcular $-2 \log \lambda$, o valor da estatística do Teste da Razão de Verossimilhanças, e obtivemos um p-valor menor que 0,003, uma evidência muito forte contra a hipótese nula. Em seguida consideramos para os dados um modelo de mistura com $g = 3$ grupos. Não serão apresentados os valores das estimativas, mas o valor da função de log-verossimilhança em sua estimativa de máxima verossimilhança foi igual a $-364,631$. O valor de $-2 \log \lambda$ para testar a hipótese de $g = 2$ contra $g = 3$ grupos foi igual a 5,69342 e o p-valor obtido com 350 amostras geradas por Bootstrap foi igual a 0,99. O procedimento realizado foi análogo ao descrito anteriormente. Portanto, não rejeitamos a hipótese de que os dados provêm de um modelo de mistura composto de $g = 2$ grupos. As probabilidades de pertencer aos grupos 1 e 2 para cada Unidade da Federação, denotados por $\hat{P}(x_j \in G_i)$, $i = 1, 2$, são dados na Tabela 9. Estes valores são as estimativas dos z_{ij} , que são os valores $b_{ij}^{*(k+1)}$, $i = 1, 2$ obtidos na iteração em que houve a convergência para os valores dos estimadores de máxima verossimilhança, obtidos na etapa E-STEP do algoritmo E.M. Veja a seção 3.3.

Portanto as Unidades da Federação classificadas com baixo número de mortes causadas por Tuberculose no ano de 2015 foram: Rondônia, Roraima, Tocantins, Piauí, Minas Gerais, Paraná, Santa Catarina, Goiás e Distrito Federal. Consequentemente, como não rejeitamos a hipótese de que o modelo é composto de 2 grupos, todas as outras 18 Unidades da Federação foram classificadas com número alto de mortes causadas por Tuberculose no ano de 2015.

Tabela 9 – Probabilidades estimadas de pertencer aos grupos 1 e 2 para cada Unidade da Federação.

Unidade da Federação	Grupo 1 $\hat{\mu}_1 = 2.25483 \times 10^{-5}$ $\hat{P}(x_j \in G_1)$	Grupo 2 $\hat{\mu}_2 = 2.70048 \times 10^{-5}$ $\hat{P}(x_j \in G_2)$
Rondônia	0.95789	0.04211
Acre	0.00214	0.99786
Amazonas	0.00000	1
Roraima	0.98012	0.01988
Pará	0.00000	1
Amapá	0.23987	0.76013
Tocantins	1.00000	0
Maranhão	0.00000	1
Piauí	0.99998	0.00002
Ceará	0.00000	1
Rio Grande do Norte	0.00740	0.9926
Paraíba	0.00000	1
Pernambuco	0.00000	1
Alagoas	0.00000	1
Sergipe	0.00109	0.99891
Bahia	0.00000	1
Minas Gerais	1.00000	0
Espírito Santo	0.00030	0.9997
Rio de Janeiro	0.00000	1
São Paulo	0.00000	1
Paraná	1.00000	0
Santa Catarina	1.00000	0
Rio Grande do Sul	0.00000	1
Mato Grosso do Sul	0.05067	0.94933
Mato Grosso	0.00000	1
Goiás	1.00000	0
Distrito Federal	1.00000	0

No APÊNDICE A encontram-se os códigos utilizados no software R para realizar os cálculos referentes a essa aplicação.

4.1.2 Identificação de outliers

Considere o conjunto de dados apresentado na Tabela 10 que pode ser encontrado em (WEISBERG, 2005). O objetivo deste exemplo é identificar outliers para o modelo de regressão linear da variável *Lpress*, definida como sendo $100 \times$ pressão atmosférica em *inHg* (polegadas de mercúrio), sobre a variável *Temp* que é a temperatura em graus $^{\circ}F$ de ebulição da água em 17 localidades dos Alpes e Escócia. A abordagem que está sendo seguida é a mesma de (AITKIN; WILSON, 1980). No entanto, neste exemplo, o procedimento a ser realizado é determinar se o conjunto de dados é proveniente de uma única população ou grupo em que todas as observações do conjunto de dados sejam adequados ao modelo de regressão linear ou se é proveniente de uma mistura de 2 grupos com distribuição Normal com mesma variância para os 2 grupos, um grupo de outliers e outro de observações que se ajustam ao modelo de regressão. Será enfatizado o caso da regressão linear simples neste texto, mas essa abordagem também pode ser aplicada no caso da regressão linear múltipla.

Para o primeiro grupo, pressupomos que y_j , a variável dependente, tem densidade dada por:

$$f_1(y_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \mu)^2}{2\sigma^2} \right\} \quad (4.7)$$

Já no segundo grupo pressupomos que cada observação y_j tem densidade dada por:

$$f_2(y_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} \right\} \quad (4.8)$$

A variável x_j é a variável independente a ser conhecida. Logo, o modelo de mistura considerado tem densidade dada por:

$$g(y_j) = \lambda_1 \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \mu)^2}{2\sigma^2} \right\} + \lambda_2 \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} \right\} \quad (4.9)$$

de forma que $\lambda_2 = 1 - \lambda_1$.

Considerando o modelo de mistura acima, obtemos a estimativa de máxima verossimilhança de $\psi = (\lambda_1, \mu, \beta_0, \beta_1, \sigma^2)$ utilizando o algoritmo E.M.. A função de log-verossimilhança completa de ψ é dada por:

$$l_C(\psi) = \sum_{j=1}^n \left[\sum_{i=1}^2 z_{ij} \log \lambda_i + z_{1j} \left(-\frac{(y_j - \mu)^2}{2\sigma^2} - \frac{1}{2} \log(2\pi) - \log \sigma \right) + z_{2j} \left(-\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} - \frac{1}{2} \log(2\pi) - \log \sigma \right) \right], \quad (4.10)$$

em que os valores de y_j e x_j , $j = 1, \dots, n$ precisam ser conhecidos e estão fixos. Logo, as estimativas dos parâmetros na iteração $(k+1)$ são dadas por

$$b_{ij}^{*(k+1)} = \frac{\lambda_i^{(k)} f_i(x_j)}{\sum_{t=1}^g \lambda_t^{(k)} f_t(x_j)} \quad (4.11)$$

na etapa E-STEP. Lembrando que $b_{2j}^{*(k+1)} = 1 - b_{1j}^{*(k+1)}$.

$$\lambda_i^{(k+1)} = \frac{\sum_{j=1}^n b_{ij}^{*(k+1)}}{n}, \quad (4.12)$$

$$\mu^{(k+1)} = \frac{\sum_{j=1}^n b_{1j}^{*(k+1)} y_j}{\sum_{j=1}^n b_{1j}^{*(k+1)}}, \quad (4.13)$$

$$\beta^{(k+1)} = (X' B_2 X)^{-1} X' B_2 y, \quad (4.14)$$

e

$$\sigma^{2(k+1)} = \frac{\sum_{j=1}^n b_{1j}^{*(k+1)} (y_j - \mu)^2 + \sum_{j=1}^n b_{2j}^{*(k+1)} (y_j - \beta_0 - \beta_1 x_j)^2}{n} \quad (4.15)$$

na etapa M-STEP.

Para os coeficientes do modelo de regressão adotamos as notações $\beta^{(k+1)} = (\beta_0^{(k+1)}, \beta_1^{(k+1)})'$, X uma matriz cuja primeira coluna é um vetor de números iguais a 1 e a segunda coluna é o vetor de variáveis independentes $x = (x_1, \dots, x_n)'$. Por último, a matriz B_2 é uma matriz diagonal cuja entrada da linha e coluna j , $j = 1, 2, \dots, n$ é igual a $b_{2j}^{*(k+1)} = 1 - b_{1j}^{*(k+1)}$. A matriz B_2 deve ser atualizada a cada iteração.

Usamos a função de log-verossimilhança $l(\mu, \beta_0, \beta_1, \sigma^2; x)$ para realizar o Teste da Razão de Verossimilhanças. Esta é dada por:

$$l(\mu, \beta_0, \beta_1, \sigma^2; x) = \sum_{j=1}^n \log \left[\lambda_1 \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \mu)^2}{2\sigma^2} \right\} + \lambda_2 \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} \right\} \right] \quad (4.16)$$

Para comparar o modelo de mistura com 2 grupos contra um modelo simples em que há somente um grupo ou população, consideramos o grupo com densidade

$$f(y_j) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} \right\}. \quad (4.17)$$

Os estimadores de máxima verossimilhança dos parâmetros β_0 , β_1 e σ^2 são os mesmos obtidos por Mínimos Quadrados Ordinários. A função de log-verossimilhança neste caso é dada por

$$l(\beta_0, \beta_1, \sigma^2; x) = - \sum_{j=1}^n \left[\frac{(y_j - \beta_0 - \beta_1 x_j)^2}{2\sigma^2} + \frac{1}{2} \log(2\pi) + \log \sigma \right]. \quad (4.18)$$

Tabela 10 – Conjunto de dados Forbes

Observação	Lpress	Temp
1	131.79	194.5
2	131.79	194.3
3	135.02	197.9
4	135.55	198.4
5	136.46	199.4
6	136.83	199.9
7	137.82	200.9
8	138.00	201.1
9	138.06	201.4
10	138.04	201.3
11	140.04	203.6
12	142.44	204.6
13	145.47	209.5
14	144.34	208.6
15	146.30	210.7
16	147.54	211.9
17	147.80	212.2

Agora vamos utilizar o conjunto de dados da Tabela 10 para identificar os outliers. Os y_j são representados pela variável Lpress e os x_j são representados pela variável Temp. O modelo de regressão para os dados é

$$Lpress = \beta_0 + \beta_1 Temp \quad (4.19)$$

Em seguida, obtemos o seguinte modelo ajustado:

$$L\hat{press} = -42.13778 + 0.89549 Temp. \quad (4.20)$$

Consequentemente, o resíduo \hat{r}_j da observação y_j é dado por $\hat{r}_j = Lpress - L\hat{press}$.

Os resíduos e valores preditos do modelo (4.20) são apresentados na Tabela 11 e o gráfico entre os valores preditos $L\hat{pres} = \hat{\beta}_0 + \hat{\beta}_1 Temp$ e os resíduos é apresentado na Figura 5. Veja que o resíduo para a observação $y_{12} = 142,44$ é distinto dos demais, $\hat{r}_{12} = 1,35978$. Esse resíduo é o maior em valor absoluto e também um outlier. Por isso, é pertinente supor que a observação $y_{12} = 142,44$ pertence a um grupo separado das demais

observações. Isto é, y_1 pertence ao grupo das observações que não se ajustam ao modelo de regressão ou grupo dos outliers.

Tabela 11 – Valores preditos e resíduos do modelo (4.20).

Observação	Preditos	Resíduos
1	132.0357	-0.24574
2	131.8566	-0.06664
3	135.0804	-0.06042
4	135.5282	0.02184
5	136.4237	0.03634
6	136.8714	-0.04140
7	137.7669	0.05310
8	137.9460	0.05400
9	138.2146	-0.15465
10	138.1251	-0.08510
11	140.1847	-0.14473
12	141.0802	1.35978
13	145.4681	0.00186
14	144.6622	-0.32220
15	146.5427	-0.24274
16	147.6173	-0.07733
17	147.8860	-0.08598

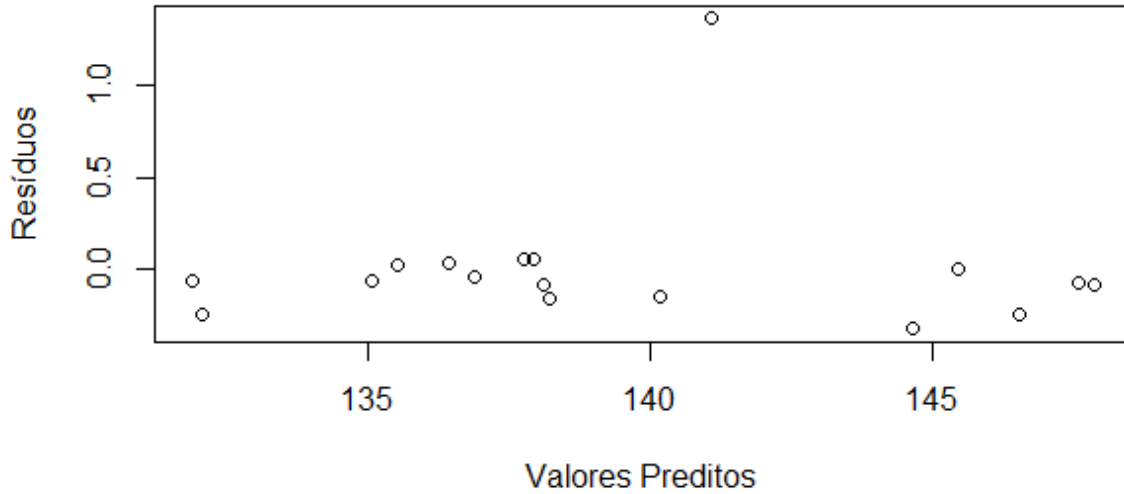


Figura 5 – Gráfico entre os valores preditos e resíduos do modelo (4.20).

Para o modelo de mistura com $g = 2$ grupos, tomando como valores iniciais $\lambda_1 = 0, 5$, $\mu_1 = y_{12} = 142,44$, $\beta = (X'B_2X)^{-1}X'B_2y = (-42, 13778; 0, 89549)'$, em que B_2 é a matriz dos $b_{2j}^{*(1)}$, $j = 1, \dots, 17$ na diagonal, obtidos na primeira iteração, e $\sigma^2 = \frac{\sum_{j=1}^n \hat{r}_j^2}{17} = 0, 12676$, os estimadores de máxima verossimilhança dos correspondentes parâmetros convergem para $\lambda_1 = 0, 05882$, $\mu_1 = 142, 44$, $\hat{\beta} = (-42, 13778; 0, 89549)'$ e $\hat{\sigma}^2 = 0, 018$. Considerando estas estimativas, o valor da função de log-verossimilhança é 6, 2243. Em comparação com um modelo simples em que há somente o grupo do modelo de regressão, o valor da função de log-verossimilhança é -6, 56558. Portanto, o valor de $-2 \log \lambda$ é 25, 57975. O p-valor obtido com Bootstrap gerando 350 amostras para o cálculo de $-2 \log \lambda$ foi menor que 0, 003. Logo, há uma evidência muito forte contra H_0 . Portanto, rejeitamos a hipótese de que o modelo de mistura para as observações é composto de apenas 1 grupo em favor da hipótese de que é composto de 2 grupos. Olhando as probabilidades das observações pertencerem ao grupo 1, o qual chamamos de Grupo dos outliers, apenas a observação y_{12} tem probabilidade maior que 0, 00001 de pertencer a esse grupo. A sua probabilidade de pertencer a este grupo é 1. Portanto, seguindo esta abordagem, há uma indicação muito forte de que a observação $y_{12} = 142, 44$ é um outlier, o que já esperávamos.

No APÊNDICE B encontram-se os códigos utilizados no software R para realizar os cálculos referentes a essa aplicação.

Conclusão

Vimos que os estimadores de máxima verossimilhança têm uma sólida teoria de inferência estatística e isso o torna o método mais adequado para se utilizar em problemas de estimação e inferência. No entanto, aplicar a inferência convencional quando as condições de regularidade não estão satisfeitas não é adequado. A distribuição dos estimadores de máxima verossimilhança, neste caso, não pode ser aproximada pela distribuição Normal e consequentemente os intervalos de confiança e testes de Wald e Razão de Verossimilhanças baseados na distribuição Normal não são válidos. Em casos especiais, a distribuição dos estimadores de máxima verossimilhança pode ser conhecida ou aproximada de alguma forma, restando apenas aplicar técnicas de reamostragem como o Bootstrap nos demais casos. Foram apresentadas aplicações em casos cujas condições de regularidade não estavam satisfeitas e o Bootstrap foi aplicado justamente por ser uma solução simples de se trabalhar e que funciona bem na maioria dos casos. Portanto, esperamos que a inferência sem a distribuição Normal com os estimadores de máxima verossimilhança se torne mais conveniente de se trabalhar, aplicando o Bootstrap e através do desenvolvimento uma alternativa mais simples, do ponto de vista prático, e mais abrangente como a teoria padrão que utiliza a aproximação com a distribuição Normal.

APÊNDICE A - Códigos de R da subseção

4.1.1

Abaixo encontram-se os códigos utilizados no software R para realizar os cálculos referentes à aplicação de classificação do número de mortes por Tuberculose na subseção 4.1.1.

```

1 ##### Classificacao do numero de mortes por Tuberculose
2
3
4 ##### Importacao dos dados
5
6 x #### vetor com numero de mortes causadas por tuberculose segundo a Unidade da Federacao
7 M #### vetor com numero de habitantes segundo a Unidade da Federacao
8
9 x = c(27,20,125,4,208,14,9,150,44,208,65,88,418,76,46,411,237,77,832,872,
10       126,50,294,49,71,63,26)
11 M=c(1768162,803503,3938325,505674,8175156,766692,1515134,6904298,3203248,8905190,
12      3442158,3972175,9345638,3340528,2242948,15203851,20869033,3929925,16550009,44396460,
13      11163023,6819124,11247923,2651222,3265468,6610683,2914830)
14
15
16 ##### calculos para o Modelo de Mistura com 2 grupos
17 #### l1 e a funcao de log-verossimilhanca dos parametros lambda_1, mi_1 e mi_2. Esta
18     funcao
19 #### tem como argumento o vetor (psi) composto dos parametros lambda_1, mi_1 e mi_2 e o
20     vetor
21 #### de obsevacoes da amostra (x). O vetor M com os valores da populacao estao fixos.
22 l1 = function(psi,x)
23 {
24   p = psi[1]
25   m1 = psi[2]
26   m2 = psi[3]
27   sum(log(p*dpois(x,m1*M)+(1-p)*dpois(x,m2*M)))
28 }
29
30 ##### le2 e uma funcao que retorna os valores dos estimadores de maxima verossimilhanca de
31     lambda_1, mi_1 e mi_2. A saida e um vetor. Os argumentos da funcao sao o vetor de
32     obsevacoes
33     da amostra (x) e o vetor (M) do numero de habitantes das Unidades da Federacao.
34
35 le2 = function(x,M)
36 {
37   #### Valores iniciais
38   p = 0.2 # estimador lambda_1
39   m1 = median(x)*27/sum(M) # estim. mi_1
40   m2 = mean(x)*27/sum(M) # estim. mi_2
41   psi1 = c(p,m1,m2)
42
43   #### Algoritmo E.M.
44   i = 1
45   e = 1
46   while(e > 0.000001)

```

```

44 {
45   ### E-STEP
46   psi0 = psi1
47   b = p*dpois(x,m1*M)/(p*dpois(x,m1*M)+(1-p)*dpois(x,m2*M)) # estim. [b_1j*]^(k+1)
48   ### M-STEP
49   p = sum(b)/27
50   m1 = sum(b*x)/sum(b*M)
51   m2 = sum((1-b)*x)/sum((1-b)*M)
52
53   psi1 = c(p,m1,m2)
54   i = i + 1
55   e = max(abs(psi1-psi0))
56 }
57 psi1
58 }
59
60 ### psi12 e o vetor com os valores dos estimadores de maxima verossimilhanca de
61 ### lambda_1, mi_1 e mi_2
62
63 psi12 = le2(x,M)
64
65 ##### Modelo simples (so ha 1 grupo)
66 ### m00 e o estimador de maxima verossimilhanca de mi_1
67 m00 = sum(x)/sum(M) # estim. mi_1
68
69 ### l0 e a funcao de log-verossimilhanca de mi_1. Os argumentos da funcao sao o valor de
70 ### mi_1 (t)
71 ### e o vetor de obsevacoes da amostra (x).
72 l0 = function(t,x)
73 {
74   sum(log(dpois(x,t*M)))
75 }
76 l0(m00,x)
77
78 ##### Teste da Razao de verossimilhanças para testar a hipotese de que o numero de mortes
79 ##### causadas por Tuberculose e o mesmo para todas as Unidades da Federacao contra a
80 ##### hipotese
81 ##### de que existem dois grupos com numero de mortes diferentes para a mesma causa. A
82 ##### primeira
83 ##### hipotese e a hipotese nula e a segunda hipotese e a hipotese alternativa.
84
85 ### Estimadores de maxima verossimilhanca de acordo com a hipotese nula e de acordo com a
86 ### hipo_
87 ### tese alternativa, respectivamente.
88 psi12 = le2(x,M)
89 m00 = sum(x)/sum(M)
90
91 ### lrt e o vetor com os valores da estatistica do Teste da Razao de verossimilhanças
92 ### para 350
93 ### amostras geradas sob a hipotese nula determinada pelo valor estimador de
94 ### maximaverossimilhanca
95 ### m00 e mais a amostra original dos dados, totalizando 351 valores.
96 lrt = rep(0,351)
97 ### lrt[351] e o valor da estatistica do Teste da Razao de verossimilhanças considerando
98 ### a amostra
99 ### original.
100 lrt[351] = -2*(l0(m00,x)-l1(psi12,x))
101 for(k in 1:350)
102 {
103   u = runif(27,0,1)

```



```

97 y = qpois(u,m0*M)
98 psi12 = le2(y,M)
99 m0 = sum(y)/sum(M)
100 lrt[k] = -2*(l0(m0,y)-l1(psi12,y))
101 }
102 ### calculo do p-valor do teste utilizando o Metodo do Percentil.
103 length(lrt[lrt>=lrt[351]])/350
104
105
106 ##### calculos para o Modelo de Mistura com 3 grupos
107
108 ### l3 e a funcao de log-verossimilhanca dos parametros lambda_1, mi_1, mi_2 e mi_3. Esta
      funcao
109 ### tem como argumento o vetor (psi) composto dos parametros lambda_1, mi_1, mi_2 e mi_3
      e o vetor
110 ### de obsevacoes da amostra (x). O vetor M com os valores da populacao estao fixos.
111 l3 = function(psi,x)
112 {
113   p1 = psi[1]
114   p2 = psi[2]
115   m1 = psi[3]
116   m2 = psi[4]
117   m3 = psi[5]
118   sum(log(p1*dpois(x,m1*M)+p2*dpois(x,m2*M)+(1-p1-p2)*dpois(x,m3*M)))
119 }
120
121 ### le3 e uma funcao que retorna os valores dos estimadores de maxima verossimilhanca de
122 ### lambda_1, mi_1, mi_2 e mi_3. A saida e um vetor. Os argumentos da funcao sao o vetor
      de obsevacoes
123 ### da amostra (x) e o vetor (M) do numero de habitantes das Unidades da Federacao.
124
125 le3 = function(x,M)
126 {
127   ### Valores iniciais
128   p1 = 0.5 # estim. lambda_1
129   p2 = 0.25 # estim. lambda_2
130   m1 = quantile(x)[[2]]*27/sum(M)
131   m2 = quantile(x)[[3]]*27/sum(M)
132   m3 = quantile(x)[[4]]*27/sum(M) # estim. mi_3
133   psi1 = c(p1,p2,m1,m2,m3)
134   ### Algoritmo E.M.
135   i = 1
136   e = 1
137   while(e > 0.00001)
138   {
139     ### E-STEP
140     psi0 = psi1
141     b1 = p1*dpois(x,m1*M)/(p1*dpois(x,m1*M)+p2*dpois(x,m2*M)+(1-p1-p2)*dpois(x,m3*M))
142     b2 = p2*dpois(x,m2*M)/(p1*dpois(x,m1*M)+p2*dpois(x,m2*M)+(1-p1-p2)*dpois(x,m3*M))
143     b3 = 1-b1-b2
144     ### M-STEP
145     p1 = (1-p2)*sum(b1)/sum(1-b2)
146     p2 = (1-p1)*sum(b2)/sum(1-b1)
147     m1 = sum(b1*x)/sum(b1*M)
148     m2 = sum(b2*x)/sum(b2*M)
149     m3 = sum(b3*x)/sum(b3*M)
150
151     psi1 = c(p1,p2,m1,m2,m3)
152     i = i + 1
153     e = max(abs(psi1-psi0))

```

```

154 }
155   psil
156 }
157
158
159 ##### Teste da Razao de verossimilhanças com Bootstrap para testar a hipótese de que o
      numero de mortes
160 ##### causadas por Tuberculose provem de um modelo de mistura com 2 grupos contra a
      hipótese de que
161 ##### o numero de mortes provem de um modelo de mistura com 3 grupos. A primeira
162 ##### hipótese é a hipótese nula e a segunda hipótese é a hipótese alternativa.
163
164 ### Estimadores de máxima verossimilhança de acordo com a hipótese nula e de acordo com a
      hipo_
165 ### tese alternativa, respectivamente.
166 psi12 = le2(x,M)
167 psi13 = le3(x,M)
168
169 ### lrt é o vetor com os valores da estatística do Teste da Razao de verossimilhanças
      para 350
170 ### amostras geradas sob a hipótese nula determinada pelo valor do estimador de máxima
      verossimilhança
171 ### psi12 é mais a amostra original dos dados, totalizando 351 valores.
172 lrt = rep(0,351)
173 lrt[351] = -2*(ll(psi12,x)-l3(psi13,x))
174 ### Geracao de amostras e calculo das estatísticas do Teste da Razao de verossimilhanças
175 for(k in 1:350)
176 {
177   ### Geracao de amostra aleatoria de acordo com o metodo da Transformacao Integral
178   u = runif(27)
179   y = rmultinom(27,1,c(psi12[1],1-psi12[1]))[1,]*qpois(u,psi12[2]*M)+
      rmultinom(27,1,c(psi12[1],1-psi12[1]))[2,]*qpois(u,psi12[3]*M)
180   psi120 = le2(y,M)
181   psi131 = le3(y,M)
182   lrt[k] = -2*(ll(psi120,y)-l3(psi131,y))
183 }
184
185 ### calculo do p-valor do teste utilizando o Metodo do Percentil.
186 length(lrt[lrt>=lrt[351]])/350

```

Listing 4.1 – Código fonte em R

APÊNDICE B - Códigos de R da subseção

4.1.2

Abaixo encontram-se os códigos utilizados no software R para realizar os cálculos referentes à aplicação de Identificação de Outliers na subseção 4.1.2.

```

1 ##### Identificacao de Outliers
2
3 ##### Dados
4 ##### O vetor y contem as observacoes da variavel Lpress e o vetor x contem as observacoes
5 da
6 ##### variavel Temp.
7
8 y = c
9   (131.79,131.79,135.02,135.55,136.46,136.83,137.82,138.00,138.06,138.04,140.04,142.44,
10    145.47,144.34,146.30,147.54,147.80)
11
12 x = c
13   (194.5,194.3,197.9,198.4,199.4,199.9,200.9,201.1,201.4,201.3,203.6,204.6,209.5,208.6,
14    210.7,211.9,212.2)
15
16 ##### Analise do modelo de regressao de Lpress sobre Temp
17 modelo = lm(y~x)
18 summary(modelo)
19 residuos = modelo$residuals
20 preditos = modelo$fitted.values
21 plot(preditos, residuos, xlab='Valores Preditos', ylab='Residuos')
22 plot(modelo)
23
24 ##### Modelo simples
25 ##### O modelo simples supoe que o conjunto de dados da regressao provem de um unico grupo
26 ##### em que nao ha ouliers, ou seja, todas as observacoes se adequam ao modelo de
27 regressao.
28
29 ##### A funcao mml retorna os valores dos estimadores de maxima verossimilhanca de beta_0
30 beta_1 e sigma^2
31 ##### em um vetor. Ela tem como argumento o vetor de observacoes da variavel dependente ou
32 ##### regressora (y).
33
34 mml = function(y)
35 {
36   b = lm(y~x)$coefficients
37   s2 = sum(modelo$residuals^2)/length(y)
38   psi = c(b,s2)
39   psi
40 }
41
42 ##### Vetor de estimativas de maxima verossimilhanca de beta_0, beta_1 e sigma^2
43 psi11 = mml(y)
44
45 ##### l11 e a funcao de verossimilhanca de beta_0, beta_1 e sigma^2 sob o modelo simples.
46 Ela tem
47 ##### como argumentos os vetores (psi), vetor de parametros com beta_0, beta_1 e sigma^2, e

```

```

      o
42 ##### vetor (y).
43
44 l11 = function(psi,y)
45 {
46   be0 = psi[1]
47   be1 = psi[2]
48   s2 = psi[3]
49   sum(log(dnorm(y,be0+be1*x,sqrt(s2))))
50 }
51 l11(psi11,y)
52
53 ##### Modelo de mistura com 2 grupos e variancias iguais
54 ##### O modelo de mistura com 2 grupos supoe que um grupo e referente aos outliers e o
      outro grupo e
55 ##### referente as observacoes que se ajustam ao modelo de regressao. Os dois grupos tem
      funcao
56 ##### de densidade Normal e as variancias sao iguais.
57
58 ### A funcao mm2l retorna os estimadores de maxima verossimilhanca de lambda 1, mu, beta_
      0,
59 ### beta_1 e de sigma^2, sob o modelo de mistura com 2 grupos. Ela tem como argumento o
      vetor (y).
60
61 mm2l = function(y)
62 {
63   ##### Valores iniciais
64   n = length(y)
65   p = 0.5 # estimativa de lambda_1
66   m1 = 142.44 # estim. de mu
67   be0 = -42.13777928 # estim. de beta_0
68   be1 = 0.89549367 # estim. de beta_1
69   s2 = var(lm(y~x)$residuals)*(n-1)/n # estim. de sigma^2
70   psi1 = c(p,m1,be0,be1,s2) # vetor com os valores dos parametros
71   i = 1
72   e = 1
73   while(e > 0.00001)
74   {
75     ##### Algoritmo E.M.
76     psi0 = psi1
77     ### E-STEP
78     ## estim. [b_1j^*]^(k+1)
79     b1= p*dnorm(y,m1,sqrt(s2))/(p*dnorm(y,m1,sqrt(s2))+(1-p)*dnorm(y,be0+be1*x,sqrt(s2)))
80     ## estim. [b_2j^*]^(k+1)
81     b2 = 1 - b1
82     ### M-STEP
83     p = sum(b1)/n
84     m1 = sum(b1*y)/sum(b1)
85     lr = function(be)
86     {
87       sum(b2*(y-be[0]-be[1]*x)^2)
88     }
89     be0 = nlm(lr,c(be0,be1))$estimate[1]
90     be1 = nlm(lr,c(be0,be1))$estimate[2]
91     s2 = sum(b1*(y-m1)^2+b2*(y-be0-be1*x)^2)/n
92     ##### Finalizando a itereracao
93     psi1 = c(p,m1,be0,be1,s2)
94     e = max(abs(psi1-psi0))
95     i = i + 1
96   }

```

```

97 |   psi1
98 | }
99 |
100 | psi21 = mm21(y)
101 |
102 | #### A funcao l21 e a funcao de log-verossimilhanca de lambda_1, mu, beta_0, beta_1 e
      sigma^2.
103 | #### Ela tem como argumento o vetor de parametros (psi) e o vetor (y).
104 |
105 | l21 = function(psi1,y)
106 | {
107 |   p = psi1[1]
108 |   m1 = psi1[2]
109 |   be0 = psi1[3]
110 |   be1 = psi1[4]
111 |   s2 = psi1[5]
112 |   sum(log(p*dnorm(y,m1,sqrt(s2))+(1-p)*dnorm(y,be0+be1*x,sqrt(s2))))
113 | }
114 | l21(psi21,y)
115 |
116 |
117 | ##### Teste da razao de verossimilhanças com Bootstrap para testar a hipotese de que os
      dados
118 | ##### sao provenientes de uma unica populacao ou grupo contra a hipotese de que os dados
      sao
119 | ##### provenientes de um modelo de mistura com 2 grupos. A primeira hipotese e a hipotese
      nula e a
120 | ##### segunda hipotese e a hipotese alternativa.
121 |
122 | #### lrt e o vetor com 351 valores da estatistica do Teste da razao de verossimilhanças
123 | #### referentes a 350 amostras geradas sob a hipotese nula, determinada pelas estimativas
      de
124 | #### maxima verossimilhança de beta_0, beta_1 e sigma_2, e mais a amostra original.
125 |
126 | lrt = rep(0,351)
127 | #### lrt[351] e o valor da estatistica do Teste da razao de verossimilhanças considerando
      a
128 | #### amostra original.
129 | lrt[351] = -2*(l11(psi11,y)-l21(psi21,y))
130 |
131 | #### Geracao de amostras e calculo das estatisticas do Teste da razao de verossimilhanças
132 | for(k in 1:350)
133 | {
134 |   n = length(y)
135 |   u = psi11[1] + psi11[2]*x + rnorm(n,0,sqrt(psi11[3]))
136 |   par0 = mml(u)
137 |   par1 = mm21(u)
138 |   lrt[k] = -2*(l11(par0,u)-l21(par1,u))
139 | }
140 | #### p-valor do teste utilizando o Metodo do Percentil
141 | length(lrt[lrt>=lrt[351]])/351

```

Listing 4.2 – Código fonte em R

Referências

- AITKIN, M.; WILSON, G. T. Mixture models, outliers, and the em algorithm. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 22, n. 3, p. 325–331, 1980. ISSN 00401706. Disponível em: <<http://www.jstor.org/stable/1268316>>. Citado na página 53.
- BOLFARINE, H.; SANDOVAL, M. *Introdução à Inferência Estatística*. 2. ed. São Paulo: SBM, 2010. Citado 2 vezes nas páginas 26 e 27.
- BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 8. ed. São Paulo: Saraiva, 2013. Citado na página 26.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. [S.l.]: Duxbury Press, 2001. Citado na página 27.
- CHERNICK, M. R. *Bootstrap Methods: A Guide for Practitioners and Researchers*. 2. ed. Hoboken, N.J: Wiley-Interscience, 2007. Citado na página 33.
- LAHIRI, S. N. *Resampling Methods for Dependent Data*. 1. ed. [S.l.]: Springer - Verlag New York, 2003. Citado na página 33.
- MCCULLOCH, C. E.; SEARLE, S. R. *Generalized, Linear, and Mixed Models*. 1. ed. [S.l.]: Wiley-Interscience, 2001. Nenhuma citação no texto.
- MCLACHLAN, G. J.; BASFORD, K. E. *Mixture Models: Inference and Applications to Clustering*. [S.l.]: CRC Press, 1987. Citado 3 vezes nas páginas 36, 41 e 43.
- MCLACHLAN, G. J.; KRISHNAN, T. *The E.M. Algorithm and Extensions*. 2. ed. Hoboken, N.J: Wiley-Interscience, 2008. Citado na página 37.
- MILLER, R. B. *Maximum likelihood estimation and inference : with examples in R, SAS, and ADMB*. 1. ed. [S.l.]: Wiley, 2011. Citado 2 vezes nas páginas 27 e 30.
- RATHIE, P.; ZÖRNIG, P. *Teoria da Probabilidade*. 1. ed. Brasília: EDU - UNB, 2012. Citado na página 26.
- SYMONS, M. J.; GRIMSON, R. C.; YUAN, Y. C. Clustering of rare events. *Biometrics*, [Wiley, International Biometric Society], v. 39, n. 1, p. 193–205, 1983. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2530819>>. Citado na página 47.
- WEISBERG, S. *Applied Linear Regression*. 3. ed. Hoboken, N.J: Wiley-Interscience, 2005. Citado na página 53.